



## NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS

The invention relates to new sequences of hepatitis C virus (HCV) genotypes and  
5 their use as prophylactic, therapeutic and diagnostic agents.

The present invention relates to new genomic nucleotide sequences and amino acid  
sequences corresponding to the coding region of these genomes. The invention relates to  
new HCV types and subtypes sequences which are different from the known HCV types  
and subtypes sequences. More particularly, the present invention relates to new HCV type 7  
10 sequences, new HCV type 9 sequences, new HCV types 10 and new HCV type 11  
sequences. Also the present invention relates to new HCV type 1 sequences of subtypes  
1d, 1e, 1f and 1g; new HCV type 2 sequences of subtypes 2e, 2f, 2g, 2h, 2i, 2k and 2l; new  
HCV type 3 sequences of subtype 3g, new HCV type 4 sequences of subtypes 4k, 4l and  
4m; a process for preparing them, and their use for diagnosis, prophylaxis and therapy.

15 The technical problem underlying the present invention is to provide new HCV  
sequences from untill now unknown HCV types and/or subtypes. More particularly, the  
present invention provides new type-specific sequences of the Core, the E1 and the NS5  
regions of new HCV types 7, 9, 10 and 11, as well as of new variants (subtypes) of HCV  
types 1, 2, 3 and 4. These new HCV sequences are useful to diagnose the presence of  
20 HCV type 1, and/or type 2, and/or type 3, and/or type 4, and/or type 7, and/or type 9, and/or  
type 10, and/or type 11 genotypes or serotypes in a biological sample. Moreover, the  
availability of these new type-specific sequences can increase the overall sensitivity of HCV  
detection and should also prove to be useful for prophylactic and therapeutic purposes.

Hepatitis C viruses (HCV) have been found to be the major cause of non-A, non-B  
25 hepatitis. The sequences of cDNA clones covering the complete genome of several  
prototype isolates have been determined (Kato et al., 1990; Choo et al., 1991; Okamoto et  
al., 1991; Okamoto et al., 1992). Comparison of these isolates shows that the variability in  
nucleotide sequences can be used to distinguish at least 2 different genotypes, type 1  
(HCV-1 and HCV-J) and type 2 (HC-J6 and HC-J8), with an average homology of about  
30 68%. Within each type, at least two subtypes exist (e.g. represented by HCV-1 and HCV-J),  
having an average homology of about 79%. HCV genomes belonging to the same subtype  
show average homologies of more than 90% (Okamoto et al., 1992). However, the partial

204020" BET 5860

nucleotide sequence of the NS5 region of the HCV-T isolates showed at most 67% homology with the previously published sequences, indicating the existence of yet another HCV type (Mori et al., 1992). Parts of the 5' untranslated region (UR), core, NS3, and NS5 regions of this type 3 have been published, further establishing the similar evolutionary  
 5 distances between the 3 major genotypes and their subtypes (Chan et al., 1992). Type 4 was subsequently discovered (Stuyver et al., 1993b; Simmonds et al., 1993a; Bukh et al., 1993; Stuyver et al., 1994a). As well as type 5 (Stuyver et al., 1993b; Simmonds et al., 1993c; Bukh et al., 1993; Stuyver et al., 1994b), and type 6 HCV groups (Bukh et al., 1993; Simmonds et al., 1993c). An overview of the present state of the art regarding HCV  
 10 genotypes is given in Table 3. The nomenclature system proposed by the inventors of the present application has now been accepted by scientists worldwide (Simmonds et al., 1994).

The aim of the present invention is to provide new HCV nucleotide and amino acid sequences enabling the detection of HCV infection.

15 Another aim of the present invention is to provide new nucleotide and amino acid HCV sequences enabling the classification of infected biological fluids into different serological groups.

Another aim of the present invention is to provide new nucleotide and amino acid HCV sequences ameliorating the overall HCV detection rate.

20 Another aim of the present invention is to provide new HCV sequences, useful for the design of HCV prophylactic or therapeutic vaccine compositions.

Another aim of the present invention is to provide a pharmaceutical composition consisting of antibodies raised against the polypeptides encoded by these new HCV sequences, for therapy or diagnosis.

25 All the aims of the present invention are met by the following embodiments of the present invention.

The present invention relates more particularly to an HCV polynucleic acid, having a nucleotide sequence which is unique to a heretofore unidentified HCV type or subtype which is different from HCV subtypes 1a, 1b, 1c, 2a, 2b, 2c, 2d, 3a, 3b, 3c, 3d, 3e, 3f, 4a,  
 30 4b, 4c, 4d, 4e, 4f, 4g, 4h, 4i, 4j, 5a or 6a, with said HCV subtypes being classified as in Table 3 by comparison of a part of the NS5 gene nucleotide sequence spanning positions 7932 to 8271, with said amino acid numbering being shown in Table 1, and with said

polynucleic acid containing at least one nucleotide differing from said known HCV nucleotide sequences, or the complement thereof. The sequence of known HCV isolates may be found in any nucleotide sequence database known in the art (such as for instance the EMBL database).

0595413-0040

The present invention thus also relates to a polynucleic acid having a nucleotide sequence which is unique to at least one of HCV subtypes 1d, 1e, 1f, 1g, 2e, 2f, 2g, 2h, 2i, 2k, 2l, 3g, 4k, 4l, 4m, 7a, 7c or 7d, with said HCV subtypes being classified as defined above.

5 The present invention thus also relates to a polynucleic acid having a nucleotide sequence which is unique to at least one of HCV types 9, 10 or 11, with said HCV types being classified as defined above.

It is to be noted that the nucleotide(s) difference in the polynucleic acids of the invention may involve an amino acid difference in the corresponding amino acid sequences  
10 encoded by said polynucleic acids. A composition according to the present invention may contain only polynucleic acid sequences or polynucleic acid sequences mixed with any excipient known in the art of diagnosis, prophylaxis or therapy.

According to a preferred embodiment, the present invention relates to a polynucleic acid encoding an HCV polyprotein comprising in its amino acid sequence at least one of the  
15 following amino acid residues:

I15, C38, V44, A49, Q43, P49, Q55, A58, S60 or D60, E68 or V68, H70, A71 or Q71 or N71, D72, H81, H101, D106, S110, L130, I134, E135, L140, S148, T150 or E150, Q153, F155, D157, G160, E165, I169, F181, L186, T190, T192 or I192 or H192, I193, A195, S196, R197 or N197 or K197, Q199 or D199 or H199 or N199, F200 or T200, A208, I213,  
20 M216 or S216, N217 or S217 or G217 or K217, T218, I219, A222, Y223, I230, W231 or L231, S232 or H232 or A232, Q233, E235 or L235, F236 or T236, F237, L240 or M240, A242, N244, N249, I250 or K250 or R250, A252 or C252, A254, I255 or V255, D256 or M256, E257, E260 or K260, R261, V268, S272 or R272, I285, G290 or F290, A291, A293 or L293 or W293, T294 or A294, S295 or H295, K296 or E296, Y297 or M297, I299 or  
25 Y299, I300, S301, P316, S2646, A2648, G2649, A2650, V2652, Q2653, H2656 or L2656, D2657, F2659, K2663 or Q2663, A2667 or V1667, D2677, L2681, M2686 or Q2686 or E2686, A2692 or K2692, H2697, I2707, L2708 or Y2708, A2709, A2719 or M2719, F2727, T2728 or D2728, E2729, F2730 or Y2730, I2741, I2745, V2746 or E2746 or L2746 or K2746, A2748, S2749 or P2749, R2750, E2751, D2752 or N2752 or S2752 or T2752 or  
30 V2752 or I2752 or Q2752, S2753 or D2753 or G2753, D2754, A2755, L2756 or Q2756, R2757,

with said notation being composed of a letter representing the amino acid residue by its

one-letter code, and a number representing the amino acid numbering according to Kato et al. (1980), as shown in Table 1,

or a part of said polynucleic acid which is unique to at least one of the HCV subtypes or types as defined in Table 5, and which contains at least one nucleotide differing from known  
5 HCV nucleotide sequences, or the complement thereof.

Each of the above-mentioned residues can be found in Figures 2, 4 or 6 showing the new amino acid sequences of the present invention aligned with known sequences of other types or subtypes of HCV for the Core/E1 region.

According to another preferred embodiment, the present invention relates to a  
10 polynucleic acid encoding a HCV polyprotein comprising in its amino acid sequence at least one amino acid sequence chosen from the following list:

- |    |  |                         |
|----|--|-------------------------|
|    | ARQSDGRSWAQ or ARRSEGRSWAQ as for subtype 1d   | (SEQ ID NO 107 and 108) |
|    | ERRPEGRSWAQ as for subtype 1e                  | (SEQ ID NO 109)         |
| 15 | ARRPEGRSWAQ as for subtype 1f                  | (SEQ ID NO 110)         |
|    | DRRTTGKSWGR as for subtype 2k                  | (SEQ ID NO 111)         |
|    | DRRATGRSWGR as for subtype 2e                  | (SEQ ID NO 112)         |
|    | DRRATGKSWGR as for subtype 2f                  | (SEQ ID NO 113)         |
|    | VRQPTGRSWGQ as for type 9                      | (SEQ ID NO 114)         |
| 20 | VRHQTGRTWAQ as for subtype 7a and 7c           | (SEQ ID NO 115)         |
|    | VRQNQGRTWAQ as for subtype 7d                  | (SEQ ID NO 116)         |
|    | ARRTEGRSWAQ as for type 10                     | (SEQ ID NO 117)         |
|    | VRRTTGRXXXX or VRRTTGRTWAQ as for type 11      | (SEQ ID NO 118 and 119) |
| 25 | HEVRNASGVYHV or HEVRNASGVYHL as for subtype 1d | (SEQ ID NO 120 and 121) |
|    | YEVHSTTDGYHV as for subtype 1f                 | (SEQ ID NO 122)         |
|    | VEVKNTSQAYMA as for subtype 2e                 | (SEQ ID NO 123)         |
|    | IQVKNNSHFYMA as for subtype 2f                 | (SEQ ID NO 124)         |
| 30 | VQVKNTSTMYMA as for subtype 2g                 | (SEQ ID NO 125)         |
|    | VQVKNTSHSYMV as for subtype 2h                 | (SEQ ID NO 126)         |
|    | VQVANRSGSYMV as for subtype 2i                 | (SEQ ID NO 127)         |

VEIKNTXNTYVL or VEIKNTSNTYVL as for subtype 2k (SEQ ID NO 128 and 129)

INYNRVSGIYYV or INYRNTSGIYHV or INYHNTSGIYHI or TNYRNVSGIYHV as for subtype 4k (SEQ ID NO 130, 131, 132 or

5 133)

QHYRNVSGIYHV as for subtype 4l (SEQ ID NO 134)

IQVKNASGIYHL as for type 9 (SEQ ID NO 135)

AHYTNKSGLYHL as for subtype 7c (SEQ ID NO 136)

LNANKSGLYHL as for subtype 7d (SEQ ID NO 137)

10 LEYRNASGLYMV as for type 10 (SEQ ID NO 138)

IYEMDGMIMHY or IYEMSGMILHA as for subtype 1d (SEQ ID NO 139 and 140)

VYEAKDIILHT as for subtype 1f (SEQ ID NO 141)

VWQLXDAVLHV as for subtype 2e (SEQ ID NO 142)

15 VWQLRDAVLHV as for subtype 2f (SEQ ID NO 143)

IWQMKGAVLHV as for subtype 2g (SEQ ID NO 144)

VWQLKDAVLHV as for subtype 2h (SEQ ID NO 145)

VWQLEEAVLHV as for subtype 2i (SEQ ID NO 146)

TWQLXXAVLHV as for subtype 2k (SEQ ID NO 147)

20 VYEADHHILHL or VYEADHHILAL or VFEADHHILHL as for subtype 4k (SEQ ID NO 148, 149 and 150)

VYESDHHILHL as for subtype 4l (SEQ ID NO 151)

VFEAETMILHL as for type 9 (SEQ ID NO 152)

VYEAETLILHL as for subtype 7c (SEQ ID NO 153)

25 VYEANGMILHL as for subtype 7d (SEQ ID NO 154)

VYEAGDIILHL as for type 10 (SEQ ID NO 155)

VREDNHLRCWMAL or VRENNSSRCWMAL as for subtype 1d (SEQ ID NO 156 and 157)

IREGNISRCWVPL as for subtype 1f (SEQ ID NO 158)

30 ENSSGRFHCWIPV as for subtype 2e (SEQ ID NO 159)

ERSGNRTFCWTAV as for subtype 2f (SEQ ID NO 160)

ELQGNKSRWCWIPV as for subtype 2g (SEQ ID NO 162)

	ERHQNSRCWIPV as for subtype 2h	(SEQ ID NO 163)
	EWKDNTSRCWIPV as for subtype 2i	(SEQ ID NO 164)
	EREGNSSRCWIPV as for subtype 2k	(SEQ ID NO 165)
5	VREGNSRCWVAL or VRTGNQSRCWVAL or VRVGNQSSCWVAL or VRVGNQSRCWVAL or VKEGNHSRCWVAL as for subtype 4k	(SEQ ID NO 166, 167, 168 or 169)
	VKTGNTSRCWVAL as for subtype 4l	(SEQ ID NO 170)
	IKAGNESRCWLPV as for type 9	(SEQ ID NO 171)
	VKEGNQSRCWVQA as for subtype 7c	(SEQ ID NO 172)
10	VKXXNLTKCWLSA as for subtype 7d	(SEQ ID NO 173)
	VRSGNTSRCWIPV as for type 10	(SEQ ID NO 174)
	VKNASVPTAA or VKDANVPTAA as for subtype 1d	(SEQ ID NO 175 and 176)
	ARIANAPIDE as for subtype 1f	(SEQ ID NO 177)
15	VSKPGALTKG as for subtype 2e	(SEQ ID NO 178)
	VSRPGALTRG as for subtype 2f	(SEQ ID NO 179)
	VNQPGALTRG as for subtype 2g	(SEQ ID NO 180)
	VSQPGALTRG as for subtype 2h	(SEQ ID NO 181)
	VSQPGALTKG as for subtype 2i	(SEQ ID NO 182)
20	VSRPGALTEG as for subtype 2k	(SEQ ID NO 183)
	APYIGAPLES or APYTAAPLES as for subtype 4k	(SEQ ID NO 184 and 185)
	APILSAPLMS as for subtype 4l	(SEQ ID NO 186)
	VPNSSVPIHG as for type 9	(SEQ ID NO 187)
25	VPNASTPVTG as for subtype 7c	(SEQ ID NO 188)
	VQNASVSIRG as for subtype 7d	(SEQ ID NO 189)
	VKSPCAATAS as for type 10	(SEQ ID NO 190)
	SPRMHHTTQE or SPRLYHTTQE as for subtype 1d	(SEQ ID NO 191 and 192)
30	TSRRHWTVQD as for subtype 1f	(SEQ ID NO 193)
	APKRHYFVQE as for subtype 2e	(SEQ ID NO 194)
	SPQYHTFVQE as for subtype 2f	(SEQ ID NO 195)

	SPQHNNFSQD as for subtype 2g	(SEQ ID NO 196)
	SPQHNFVQD as for subtype 2h	(SEQ ID NO 197)
	SPEHHHFVQD as for subtype 2k	(SEQ ID NO 198)
5	RPRRHWTQD or RPRRHWTQD or QPRRHWTQD or RPRRHWTQE as for subtype 4k	(SEQ ID NO 199, 200, 201 or 202)
	QPRRHWTQD as for subtype 4l	(SEQ ID NO 203)
	RPKYHQVTQD as for type 9	(SEQ ID NO 204)
	RPRMHQVVQE as for subtype 7c	(SEQ ID NO 205)
10	RPRMYEIAQD as for subtype 7d	(SEQ ID NO 206)
	RHRQHWTQD as for type 10	(SEQ ID NO 207)

or a part of said polynucleic acid which is unique to at least one of the HCV subtypes or types as defined Table 5, and which contains at least one nucleotide differing from known HCV nucleotide sequences, or the complement thereof.

15 Using the 5' non-coding LiPA system (Stuyver et al., 1993) and a new core LiPA system including multiple probes for subtypes 1a, 1b, 1c, 2a, 2b or 2c derived from the core region (Stuyver et al., 1995), samples from the Benelux, Cameroon, France and Vietnam were selected because of their aberrant reactivities (isolates CAM1078, FR2, FR1, VN4, VN12, VN13, NE98). Some samples were, together with many other samples, sequenced  
20 as a control for typing. Sequencing results, however, indicated the discovery of new subtypes (isolates BNL1, BNL2, BNL3, FR4, BNL4, BNL5, BNL6, BNL7, BNL8, BNL9, BNL10, BNL11 and BNL12). Nucleotide sequences in the core and E1 regions which have not yet been reported before, were analyzed in the frame of the invention. Genomic  
25 sequences of subtype 1d, 1e, 1f, 1g 2e, 2f, 2g, 2h, 2i, 2k, 2l, 3g, 4k, 4l, 4m, 7a, 7c, 7d and types 9, 10 and 11 isolates are reported for the first time in the present invention. The NS5B region was also analyzed.

The term "polynucleic acid" refers to a single- stranded or double-stranded nucleic acid sequence which may contain at least 5 contiguous nucleotides in common with the complete nucleotide sequence (e.g. at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19,  
30 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 75 or more contiguous nucleotides). A polynucleic acid which is up till about 100 nucleotides in length is often also referred to as an oligonucleotide. A polynucleic acid may consist of deoxyribonucleotides or

ribonucleotides, nucleotide analogues or modified nucleotides, or may have been adapted for therapeutic purposes. A polynucleic acid may also comprise a double stranded cDNA clone which can be used for cloning purposes, or for *in vivo* therapy, or prophylaxis.

The oligonucleotides according to the present invention, used as primers or probes  
5 may also contain or consist of nucleotide analogous such as phosphorothioates (Matsukura et al., 1987), alkylphosphorates (Miller et al., 1979) or peptide nucleic acids (Nielsen et al., 1991; Nielsen et al., 1993) or may contain intercalating agents (Asseline et al., 1984).

As most other variations or modifications introduced into the original DNA sequences of the invention these variations will necessitate adaptations with respect to the  
10 conditions under which the oligonucleotide should be used to obtain the required specificity and sensitivity. However the eventual results will be essentially the same as those obtained with the unmodified oligonucleotides.

The introduction of these modifications may be advantageous in order to positively influence characteristics such as hybridization kinetics, reversibility of the hybrid-formation,  
15 biological stability of the oligonucleotide molecules, etc.

The polynucleic acids of the invention may be comprised in a composition of any kind. Said composition may be for diagnostic, therapeutic or prophylactic use.

The expression "sequences which are unique to an HCV type or subtype" refers to sequences which are not shared by any other type or subtype of HCV, and can thus be  
20 used to uniquely detect that HCV type or subtype. Sequence variability is demonstrated in the present invention between the newly found HCV types and subtypes (see Table 5) and the known HCV types and subtypes (see Table 3), and it is therefore from these regions of sequence variability in particular that type- or subtypes-specific polynucleic acids, oligonucleotides, polypeptides and peptides may be obtained. The term type- or subtypes-  
25 specific refers to the fact that a sequence is unique to that HCV type or subtype involved.

The expression "nucleotides corresponding to" refers to nucleotides which are homologous or complementary to an indicated nucleotide sequence or region within a specific HCV sequence.

The term "coding region" corresponds to the region of the HCV genome that  
30 encodes the HCV polyprotein. In fact, it comprises the complete genome with the exception of the 5' untranslated region and 3' untranslated region.

The term "HCV polyprotein" refers to the HCV polyprotein of the HCV-J isolate (Kato

et al., 1990). The adenine residue at position 330 (Kato et al., 1990) is the first residue of the ATG codon that initiates the long HCV polypeptide of 3010 amino acids in HCV-J and other type 1b isolates, and of 3011 amino acids in HCV-1 and other type 1a isolates, and of 3033 amino acids in type 2 isolates HC-J6 and HC-J8 (Okamoto et al., 1992).

5 This adenine is designated as position 1 at the nucleic acid level, and this methionine is designated as position 1 at the amino acid level, in the present invention. As type 1a isolates contain 1 extra amino acid in the NS5A region, coding sequences of type 1a and 1b have identical numbering in the Core, E1, NS3, and NS4 region, but will differ in the NS5B region as indicated in Table 1. Type 2 isolates have 4 extra amino acids in the E2  
10 region, and 17 or 18 extra amino acids in the NS5 region compared to type 1 isolates, and will differ in numbering from type 1 isolates in the NS3/4 region and NS5b regions as indicated in Table 1. Similar insertions compared with type 1 (but of a different size) can also be observed in type 3a sequences which affect the numbering of type 3a amino acids accordingly. Other insertions or deletions may be readily observed in type1, 2, 3, 4, 5, 6, 7,  
15 8, 9, 10 and 11 sequences after alignment with known HCV sequences.

TABLE 1

	Region	Positions described in the present invention*	Positions described for HCV-J (Kato et al., 1990)	Positions described for HCV-1 (Choo et al., 1991)	Positions described for HC-J6, HC-J8 (Okamoto et al., 1992)
Nucleotides	NS5B	8023/8235 7932/8271	8352/8564 8261/8600	8026/8238 7935/8274	8433/8645 8342/8681
		coding region of present invention	330/9359	1/9033	342/9439
Amino Acids	NS5B	2675/2745 2645/2757	2675/2745 2645/2757	2676/2746 2646/2758	2698/2768 2668/2780

Table 1: Comparison of the HCV nucleotide and amino acid numbering system used in the present invention (\*) with the numbering used for other prototype

isolates. For example, 8352/8564 indicates the region designated by the numbering from nucleotide 8352 to nucleotide 8564 as described by Kato et al. (1990). Since the numbering system of the present invention starts at the polyprotein initiation site, the 329 nucleotides of the 5' untranslated region described by Kato et al. (1990) have to be subtracted, and the corresponding region is numbered from nucleotide 8023 ('8352-329') to 8235 ('8564-329').

The term "genotype" as used in the present invention refers to both types and/or subtypes.

The term "HCV type" corresponds to a group of HCV isolates of which the complete genome shows more than 73% preferably more than 74% homology at the nucleic acid level, or of which the NS5 region between nucleotide positions 7932 and 8271 shows more than 75.4% homology at the nucleic acid level, or of which the complete HCV polyprotein shows more than 78% homology at the amino acid level, or of which the NS5 region between amino acids at positions 2645 and 2757 shows more than 80% homology at the amino acid level, to polyproteins of the other isolates of the group, with said numbering beginning at the first ATG codon or first methionine of the long HCV polyprotein of the HCV-J isolate (Kato et al., 1990). Isolates belonging to different types of HCV exhibit homologies, over the complete genome, of less than 74%, preferably less than 73%, at the nucleic acid level and less than 78% at the amino acid level. Isolates belonging to the same type usually show homologies of about 90 to 99% at the nucleic acid level and 95 to 96% at the amino acid level when belonging to the same subtype, and those belonging to the same type but different subtypes preferably show homologies of about 76% to 82% (more particularly of about 77% to 80%) at the nucleic acid level and 85-86% at the amino acid level.

More preferably the definition of HCV types is concluded from the classification of HCV isolates according to their nucleotide distances calculated as detailed below:

(1) based on phylogenetic analysis of nucleic acid sequences in the NS5B region between nucleotides 7935 and 8274 (Choo et al., 1991) or 8261 and 8600 (Kato et al., 1990) or 8342 and 8681 (Okamoto et al., 1991), isolates belonging to the same HCV type show nucleotide distances of less than 0.34, usually less than 0.33, and more usually of less than 0.32, and isolates belonging to the same subtype show nucleotide distances of less

than 0.135, usually of less than 0.13, and more usually of less than 0.125, usually ranging between 0.0003 and 0.1151, and consequently isolates belonging to the same type but different subtypes show nucleotide distances ranging from 0.135 to 0.34, usually ranging from 0.1384 to 0.2977, and more usually ranging from 0.15 to 0.32, and isolates belonging  
5 to different HCV types show nucleotide distances greater than 0.34, usually greater than 0.35, and more usually of greater than 0.358, more usually ranging from 0.3581 to 0.6670.

(2) based on phylogenetic analysis of nucleic acid sequences in the core/E1 region between nucleotides 378 and 957, isolates belonging to the same HCV type show nucleotide distances of less than 0.38, usually of less than 0.37, and more usually of less  
10 than 0.364, and isolates belonging to the same subtype show nucleotide distances of less than 0.17, usually of less than 0.16, and more usually of less than 0.15, more usually less than 0.135, more usually less than 0.134, and consequently isolates belonging to the same type but different subtypes show nucleotide distances ranging from 0.15 to 0.38, usually ranging from 0.16 to 0.37, and more usually ranging from 0.17 to 0.36, more usually ranging  
15 from 0.133 to 0.379, and isolates belonging to different HCV types show nucleotide distances greater than 0.34, 0.35, 0.36, usually more than 0.365, and more usually of greater than 0.37,

Table 2 : Molecular evolutionary distances

Region	Core/E1 579 bp	E1 384 bp	NS5B 340 bp	NS5B 222 bp
Isolates*	0.0017 - 0.1347 (0.0750 $\pm$ 0.0245)	0.0026 - 0.2031 (0.0969 $\pm$ 0.0289)	0.0003 - 0.1151 (0.0637 $\pm$ 0.0229)	0.000 - 0.1323 (0.0607 $\pm$ 0.0205)
Subtypes*	0.1330 - 0.3794 (0.2786 $\pm$ 0.0363)	0.1645 - 0.4869 (0.3761 $\pm$ 0.0433)	0.1384 - 0.2977 (0.2219 $\pm$ 0.0341)	0.117 - 0.3538 (0.2391 $\pm$ 0.0399)
Types*	0.3479 - 0.6306 (0.4703 $\pm$ 0.0525)	0.4309 - 0.9561 (0.6308 $\pm$ 0.0928)	0.3581 - 0.6670 (0.4994 $\pm$ 0.0495)	0.3457 - 0.7471 (0.5295 $\pm$ 0.0627)

Table 2

Figures created by the PHYLIP program DNADIST are expressed as minimum to maximum (average  $\pm$  standard deviation). Phylogenetic distances for isolates belonging to the same subtype ('isolates'), to different subtypes of the same type ('subtypes'), and to different types ('types') are given.

In a comparative phylogenetic analysis of available sequences, ranges of molecular evolutionary distances for different regions of the genome were calculated, based on 19,781 pairwise comparisons by means of the DNADIST program of the phylogeny inference package PHYLIP version 3.5c (Felsenstein, 1993). The results are shown in Table 2 and indicate that although the majority of distances obtained in each region fit with classification of a certain isolate, only the ranges obtained in the 340bp NS5B-region are non-overlapping and therefore conclusive. However, as was performed in the present invention, it is preferable to obtain sequence information from at least 2 regions before final classification of a given isolate.

Designation of a number to the different types of HCV and HCV nomenclature is based on chronological discovery of the different types. The numbering system used in the present invention might still fluctuate according to international conventions or guidelines. For example, "type 4" might be changed into "type 5" or "type 6". Also the arbitrarily chosen border distances between types and subtypes and isolates may still be subject to change according to international guidelines or conventions. Therefore types 7a, 8a, 8b, 9a may for example be designated 6b, 6c, 6d, and 6d in the future; and type 10a which shows relatedness with genotype 3 may be denoted 3g instead of 10a.

The term "subtype" corresponds to a group of HCV isolates of which the complete polyprotein shows a homology of more than 90% both at the nucleic acid and amino acid levels, or of which the NS5 region between nucleotide positions 7932 and 8271 shows a homology of more than 90% at the nucleic acid level to the corresponding parts of the genomes of the other isolates of the same group, with said numbering beginning with the adenine residue of the initiation codon of the HCV polyprotein. Isolates belonging to the same type but different subtypes of HCV show homologies of more than 74% at the nucleic acid level and of more than 78% at the amino acid level.

It is to be understood that extremely variable regions such as the E1, E2 and NS4

regions will exhibit lower homologies than the average homology of the complete genome of the polyprotein.

Using these criteria, HCV isolates can be classified into at least 11 types. Several subtypes can clearly be distinguished in types 1, 2, 3, 4 and 7 : 1a, 1b, 1c, 1d, 1e, 1f, 1g, 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h, 2i, 2k, 2l, 3a, 3b, 3c, 3d, 3f, 3g, 4a, 4b, 4c, 4d, 4e, 4f, 4g, 4h, 4i, 4j, 4k, 4l, 4m, 7a, 7c, and 7d based on homologies of the 5' UR and coding regions. An overview of most of the reported isolates and their proposed classification according to the typing system of the present invention as well as other proposed classifications is presented in Table 3.

10

15

**Table 3**

	<i>HCV CLASSIFICATION</i>				
	OKA-MOTO	MORI	CHA	NAKAO	PROTOTYPE
1a	I	I	Pt	GI	HCV-1, HCV-H, HC-J1
1b	II	II	KI	GII	HCV-J, HCV-BK, HCV-T, HC-JK1, HC-J4, HCV-CHINA
1c					HC-G9
2a	III	III	K2a	GIII	HC-J6
2b	IV	IV	K2b	GIII	HC-J8
2c					S83, ARG6, ARG8, I10, T983
2d					NE92
3a	V	V	K3	GIV	BR36, BR56, HD10, N2L1, BR33, Ta, E-b1
3b		VI	K3	GIV	HCV-TR, Tb, NE137
3c					NE48
3d					NE274
3e					NE145

3f		NE125
4a		Z4, GB809-4
4b		Z1
4c		GB116, GB358, GB215, Z6, Z7
4d		DK13
4e		GB809-2, CAM600, CAM736
4f		CAM622, CAM627
4g		GB549
4h		GB438
4i		CAR4/1205
4j		CAR1/905
5a	GV	SA3, SA4, SA1, SA7, SA11, BE95
6a		HK1, HK2, HK3, HK4, VN11

Table 3 Overview of the known HCV types and subtypes classified according to the different authors.

5

The term "complement" refers to a nucleotide sequence which is complementary to an indicated sequence and which is able to hybridize to the indicated sequences.

The composition of the invention can comprise many combinations. By way of example, the composition of the invention can comprise:

- two (or more) nucleic acids from the same region or,
- two nucleic acids (or more), respectively from different regions, for the same isolate or for different isolates,
- or nucleic acids from the same regions and from at least two different regions (for the same isolate or for different isolates).

15

The present invention relates particularly to a polynucleic acid as defined above having a sequence selected from any of SEQ ID NO 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103 to 105, or a part of said

polynucleic acid which is unique to any of the HCV subtypes or types as defined in Table 5, and which contains at least one nucleotide differing from known HCV polynucleic acids, or the complement thereof.

The present invention relates more particularly to a polynucleic acid as defined  
5 above, which codes for the 5' UR, the Core/E1, the NS4 or the NS5B region or a part thereof.

More particularly, the present invention relates to a polynucleic acid as defined above which is a cDNA sequence.

Also included within the present invention are sequence variants of the polynucleic  
10 acids as selected from any of the nucleotide sequences as given in any of the above given SEQ ID numbers with said sequence variants containing either deletion and/or insertions of one or more nucleotides, especially insertions or deletions of 1 or more codons, mainly at the extremities of oligonucleotides (either 3' or 5'), or substitutions of some non-essential nucleotides (i.e. nucleotides not essential to discriminate between different genotypes of  
15 HCV) by others (including modified nucleotides an/or inosine), for example, a type 1 or 2 sequence might be modified into a type 7 sequence by replacing some nucleotides of the type 1 or 2 sequence with type-specific nucleotides of type 7 as shown in for instance Figure 1 and 2.

Particularly preferred variant polynucleic acids of the present invention include also  
20 sequences which hybridise under stringent conditions with any of the polynucleic acid sequences of the present invention. Particularly, sequences which show a high degree of homology (similarity) to any of the polynucleic acids of the invention as described above. Particularly sequences which are at least 80%, 85%, 90%, 95% or more homologous to said polynucleic acid sequences of the invention. Preferably said sequences will have less  
25 than 20%, 15%, 10%, or 5% variation of the original nucleotides of said polynucleic acid sequence.

Polynucleic acid sequences according to the present invention which are homologous to the sequences as represented by a SEQ ID NO can be characterized and isolated according to any of the techniques known in the art, such as amplification by means  
30 of sequence-specific primers, hybridization with sequence-specific probes under more or less stringent conditions, serological screening methods or via the LiPA typing system.

Other preferred variant polynucleic acids of the present invention include sequences

which are redundant as a result of the degeneracy of the genetic code compared any of the above-given polynucleic acids of the present invention. These variant polynucleic acid sequences will thus encode the same amino acid sequence as the polynucleic acids they are derived from.

5 Also included within the scope of the present invention are 5' non-coding region sequences which can be readily obtained from type 1 subtype 1d, 1e, 1f or 1g isolates; type 2 subtype 2e, 2f, 2g, 2h, 2i, 2k or 2l isolates; type 3 subtype 3g isolates; type 4 subtype 4k, 4l or 4m isolates; type 7 subtype 7a, 7c or 7d isolates, type 9, type 10 or type 11 isolates discribed herein. Such sequences may contain type or subtype-specific motifs which can be  
10 employed for type and/or subtype-specific hybridization assays, e.g. such as described by Stuyver et al. (1993).

Polynucleic acid sequences of the genomes indicated above from regions not yet depicted in the present examples, figures and sequence listing can be obtained by any of the techniques known in the art, such as amplification techniques using suitable primers  
15 from the sequences of these new genomes given in Figure 1 of the present invention.

The present invention also relates to an oligonucleotide primer comprising part of a polynucleic acid as defined above, with said primer being able to act as a primer for specifically amplifying the nucleic acid of a certian HCV isolate belonging to the genotype from which the primer is derived.

20 The term "primer" refers to a single stranded DNA oligonucleotide sequence capable of acting as a point of initiation for synthesis of a primer extension product which is complementary to the nucleic acid strand to be copied. The length and the sequence of the primer must be such that they allow to prime the synthesis of the extension products. Preferably the primer is about 5-50 nucleotides. Specific length and sequence will depend  
25 on the complexity of the required DNA or RNA targets, as well as on the conditions of primer use such as temperature and ionic strength.

The fact that amplification primers do not have to match exactly with corresponding template sequence to warrant proper amplification is amply documented in the literature (Kwok et al., 1990).

30 The amplification method used can be either polymerase chain reaction (PCR; Saiki et al., 1988), ligase chain reaction (LCR; Landgren et al., 1988; Wu & Wallace, 1989; Barany, 1991), nucleic acid sequence-based amplification (NASBA; Guatelli et al., 1990;

Compton, 1991), transcription-based amplification system (TAS; Kwoh et al., 1989), strand displacement amplification (SDA; Duck, 1990; Walker et al., 1992) or amplification by means of Q $\beta$  replicase (Lizardi et al., 1988; Lomeli et al., 1989) or any other suitable method to amplify nucleic acid molecules using primer extension. During amplification, the amplified products can be conveniently labelled either using labelled primers or by incorporating labelled nucleotides. Labels may be isotopic ( $^{32}\text{P}$ ,  $^{35}\text{S}$ , etc.) or non-isotopic (biotin, digoxigenin, etc.). The amplification reaction is repeated between 20 and 70 times, advantageously between 25 and 45 times.

The present invention also relates to an oligonucleotide probe comprising part of a polynucleic acid as defined above, with said probe being able to act as a hybridization probe for specific detection and/or classification into types and/or subtypes of an HCV nucleic acid containing said nucleotide sequence, with said probe being possibly labelled or attached to a solid substrate.

The term "probe" refers to single stranded sequence-specific oligonucleotides which have a sequence which is complementary to the target sequence of the HCV genotype(s) to be detected.

Preferably, these probes are about 5 to 50 nucleotides long, more preferably from about 10 to 25 nucleotides.

The term "solid support" can refer to any substrate to which an oligonucleotide probe can be coupled, provided that it retains its hybridization characteristics and provided that the background level of hybridization remains low. Usually the solid substrate will be a microtiter plate, a membrane (e.g. nylon or nitrocellulose) or a microsphere (bead). Prior to application to the membrane or fixation it may be convenient to modify the nucleic acid probe in order to facilitate fixation or improve the hybridization efficiency. Such modifications may encompass homopolymer tailing, coupling with different reactive groups such as aliphatic groups,  $\text{NH}_2$  groups, SH groups, carboxylic groups, or coupling with biotin or haptens.

The present invention also relates to a diagnostic kit for use in determining the genotype of HCV, said kit comprising a primer as defined above.

The present invention also relates to a diagnostic kit for use in determining the genotype of HCV, said kit comprising a probe as defined above.

The present invention also relates to a diagnostic kit as defined above, wherein said probe(s) is(are) attached to a solid substrate.

The present invention also relates to a diagnostic kit as defined above, wherein a range of said probes is attached to specific locations on a solid substrate.

The present invention also relates to a diagnostic kit as defined above, wherein said solid support is a membrane strip and said probes are coupled to the membrane in the form of parallel lines.

The present invention also relates to a method for the detection of HCV nucleic acids present in a biological sample, comprising:

- (i) possibly extracting sample nucleic acid,
- (ii) amplifying the nucleic acid with at least one primer as defined above,
- 10 (iii) detecting the amplified nucleic acids.

The present invention also relates to a method for the detection of HCV nucleic acids present in a biological sample, comprising:

- (i) possibly extracting sample nucleic acid,
- (ii) possibly amplifying the nucleic acid with at least one primer as defined above, or
- 15 with a universal HCV primer,
- (iii) hybridizing the nucleic acids of the biological sample, possibly under denatured conditions, at appropriate conditions with one or more probes as defined above, with said probes being preferably attached to a solid substrate,
- (iv) possibly washing at appropriate conditions,
- 20 (v) detecting the hybrids formed.

The present invention also relates to a method for detecting the presence of one or more HCV genotypes present in a biological sample, comprising:

- (i) possibly extracting sample nucleic acid,
- (ii) specifically amplifying the nucleic acid with at least one primer as defined above,
- 25 (iii) detecting said amplified nucleic acids.

The present invention also relates to a method for detecting the presence of one or more HCV genotypes present in a biological sample, comprising:

- (i) possibly extracting sample nucleic acid,
- (ii) possibly amplifying the nucleic acid with at least one primer as defined above or
- 30 with a universal HCV primer,
- (iii) hybridizing the nucleic acids of the biological sample, possibly under denatured conditions, at appropriate conditions with one or more probes as defined above,

with said probes being preferably attached to a solid substrate,

- (iv) possibly washing at appropriate conditions,
  - (v) detecting the hybrids formed,
  - (vi) inferring the presence of one or more HCV genotypes present from the observed
- 5 hybridization pattern.

The present invention also relates to a method as defined above, wherein said probes are further characterized as defined above.

The present invention also relates to a method as defined above, wherein said nucleic acids are labelled during or after amplification.

- 10 Preferably, this technique could be performed in the 5' non-coding, Core or NS5B region.

The term "nucleic acid" can also be referred to as analyte strand and corresponds to a single- or double-stranded nucleic acid molecule. This analyte strand is preferentially positive- or negative stranded RNA, cDNA or amplified cDNA.

- 15 The term "biological sample" refers to any biological sample (tissue or fluid) containing HCV nucleic acid sequences and refers more particularly to blood serum or plasma samples.

The term "universal HCV primer" refers to oligonucleotide sequences complementary to any of the conserved regions of the HCV genome.

- 20 The expression "appropriate" hybridization and washing conditions are to be understood as stringent and are generally known in the art (e.g. Maniatis et al., Molecular Cloning: A Laboratory Manual, New York, Cold Spring Harbor Laboratory, 1982).

However, according to the hybridization solution (SSC, SSPE, etc.), these probes should be hybridized at their appropriate temperature in order to attain sufficient specificity.

- 25 The term "labelled" refers to the use of labelled nucleic acids. This may include the use of labelled nucleotides incorporated during the polymerase step of the amplification such as illustrated by Saiki et al. (1988) or Bej et al. (1990) or labelled primers, or by any other method known to the person skilled in the art.

- The process of the invention comprises the steps of contacting any of the probes as
- 30 defined above, with one of the following elements:

- either a biological sample in which the nucleic acids are made available for hybridization,

- or the purified nucleic acids contained in the biological sample
- or a single copy derived from the purified nucleic acids,
- or an amplified copy derived from the purified nucleic acids, with said elements or with said probes being attached to a solid substrate.

5 The expression "inferring the presence of one or more HCV genotypes present from the observed hybridization pattern" refers to the identification of the presence of HCV genomes in the sample by analyzing the pattern of binding of a panel of oligonucleotide probes. Single probes may provide useful information concerning the presence or absence of HCV genomes in a sample. On the other hand, the variation of the HCV genomes is  
 10 dispersed in nature, so rarely is any one probe able to identify uniquely a specific HCV genome. Rather, the identity of an HCV genotype may be inferred from the pattern of binding of a panel of oligonucleotide probes, which are specific for (different) segments of the different HCV genomes. Depending on the choice of these oligonucleotide probes, each known HCV genotype will correspond to a specific hybridization pattern upon use of a  
 15 specific combination of probes. Each HCV genotype will also be able to be discriminated from any other HCV genotype amplified with the same primers depending on the choice of the oligonucleotide probes. Comparison of the generated pattern of positively hybridizing probes for a sample containing one or more unknown HCV sequences to a scheme of expected hybridization patterns, allows one to clearly infer the HCV genotypes present in  
 20 said sample.

The present invention thus relates to a method as defined above, wherein one or more hybridization probes are selected from any of SEQ ID NO 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103 or 105 or  
 25 sequence variants thereof as defined above.

In order to distinguish the amplified HCV genomes from each other, the target polynucleic acids are hybridized to a set of sequence-specific DNA probes targetting HCV genotypic regions (unique regions) located in the HCV polynucleic acids.

Most of these probes target the most type- or subtype-specific regions of HCV  
 30 genotypes, but some can be caused to hybridize to more than one HCV genotype.

According to the hybridization solution (SSC, SSPE, etc.), these probes should be stringently hybridized at their appropriate temperature in order to attain sufficient specificity.

However, by slightly modifying the DNA probes, either by adding or deleting one or a few nucleotides at their extremities (either 3' or 5'), or substituting some non-essential nucleotides (i.e. nucleotides not essential to discriminate between types) by others (including modified nucleotides or inosine) these probes or variants thereof can be caused  
5 to hybridize specifically at the same hybridization conditions (i.e. the same temperature and the same hybridization solution). Also changing the amount (concentration) of probe used may be beneficial to obtain more specific hybridization results. It should be noted in this context, that probes of the same length, regardless of their GC content, will hybridize specifically at approximately the same temperature in TMAcI solutions (Jacobs et al., 1988).

10 Suitable assay methods for purposes of the present invention to detect hybrids formed between the oligonucleotide probes and the nucleic acid sequences in a sample may comprise any of the assay formats known in the art, such as the conventional dot-blot format, sandwich hybridization or reverse hybridization. For example, the detection can be accomplished using a dot blot format, the unlabelled amplified sample being bound to a  
15 membrane, the membrane being incorporated with at least one labelled probe under suitable hybridization and wash conditions, and the presence of bound probe being monitored.

An alternative and preferred method is a "reverse" dot-blot format, in which the amplified sequence contains a label. In this format, the unlabelled oligonucleotide probes  
20 are bound to a solid support and exposed to the labelled sample under appropriate stringent hybridization and subsequent washing conditions. It is to be understood that also any other assay method which relies on the formation of a hybrid between the nucleic acids of the sample and the oligonucleotide probes according to the present invention may be used.

According to an advantageous embodiment, the process of detecting one or more  
25 HCV genotypes contained in a biological sample comprises the steps of contacting amplified HCV nucleic acid copies derived from the biological sample, with oligonucleotide probes which have been immobilized as parallel lines on a solid support.

According to this advantageous method, the probes are immobilized in a Line Probe Assay (LiPA) format. This is a reverse hybridization format (Saiki et al., 1989) using  
30 membrane strips onto which several oligonucleotide probes (including negative or positive control oligonucleotides) can be conveniently applied as parallel lines.

The invention thus also relates to a solid support, preferably a membrane strip,

carrying on its surface, one or more probes as defined above, coupled to the support in the form of parallel lines.

The LiPA is a very rapid and user-friendly hybridization test. Results can be read after 4 hours. after the start of the amplification. After amplification during which usually a non-isotopic label is incorporated in the amplified product, and alkaline denaturation, the amplified product is contacted with the probes on the membrane and the hybridization is carried out for about 1 to 1,5 h hybridized polynucleic acid is detected. From the hybridization pattern generated, the HCV type can be deduced either visually, but preferably using dedicated software. The LiPA format is completely compatible with commercially available scanning devices, thus rendering automatic interpretation of the results very reliable. All those advantages make the LiPA format liable for the use of HCV detection in a routine setting. The LiPA format should be particularly advantageous for detecting the presence of different HCV genotypes.

The present invention also relates to a method for detecting and identifying novel HCV genotypes, different from the known HCV genomes, comprising the steps of:

- determining to which HCV genotype the nucleotides present in a biological sample belong, according to the process as defined above,
- in the case of observing a sample which does not generate a hybridization pattern compatible with those defined in Table 3, sequencing the portion of the HCV genome sequence corresponding to the aberrantly hybridizing probe of the new HCV genotype to be determined.

The present invention also relates to a method for preparing a polynucleic acid according to the present invention. These methods include any method known in the art for preparing polynucleic acids (e.g. the phosphodiester method for synthesizing oligonucleotides as described by Agarwal et al. 1972, Agnew. Chem. Int. Ed. Engl. 11:451, the phosphotriester method of Hsiung et al. 1979, Nucleic Acid Res. 6:1371, or the automated diethylphosphoramidite method of Baeucage et al. 1981, Tetrahedron Letters 22:1859-1862.). Alternatively, the polynucleic acids of the present invention may be isolated fragments of naturally occurring or cloned DNA or RNA. In addition, the oligonucleotides according to the present invention may be synthesized automatically on commercial instruments sold by a variety of manufacturers.

The present invention particularly also relates to a polypeptide having an amino acid

sequence encoded by a polynucleic acid as defined above, or a part thereof which is unique to at least one of the HCV subtypes or types as defined in Table 5, and which contains at least one amino acid differing from any of the known HCV types or subtypes, or an analog thereof being substantially homologous and biologically equivalent .

5       The term 'polypeptide' refers to a polymer of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or  
10 more analogues of an amino acid (including, for example, unnatural amino acids, PNA, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The term "unique" is referred above.

By "biologically equivalent" as used throughout the specification and claims, it is  
15 meant that the compositions are immunogenically equivalent to the proteins (polypeptides) or peptides of the invention as defined above and below.

By "substantially homologous" as used throughout the ensuing specification and claims to describe proteins and peptides, it is meant a degree of homology in the amino acid sequence to the proteins or peptides of the invention. Preferably the degree of homology is  
20 in excess of 90, preferably in excess of 95, with a particularly preferred group of proteins being in excess of 99 homologous with the proteins or peptides of the invention.

The term "analog" as used throughout the specification or claims to describe the proteins or peptides of the present invention, includes any protein or peptide having an amino acid residue sequence substantially identical to a sequence specifically shown herein  
25 in which one or more residues have been conservatively substituted with a biologically equivalent residue. Examples of conservative substitutions include the substitution of one-polar (hydrophobic) residue such as isoleucine, valine, leucine or methionine for another, the substitution of one polar (hydrophillic) residue for another such as between arginine and lysine, between glutamine and asparagine, between glycine and serine, the substitution of  
30 one basic residue such as lysine, arginine or histidine for another, or the substitution of one acidic residue, such as aspartic acid or glutamic acid for another. Examples of allowable mutations according to the present invention can be found in Table 4.

The phrase "conservative substitution" also includes the use of a chemically derivatized residue in place of a non-derivatized residue provided that the resulting protein or peptide is biologically equivalent to the protein or peptide of the invention.

"Chemical derivative" refers to a protein or peptide having one or more residues 5 chemically derivatized by reaction of a functional side group. Examples of such derivatized molecules, include but are not limited to, those molecules in which free amino groups have been derivatized to form amine hydrochlorides, p-toluene sulfonyl groups, carbobenzoxy groups, t-butyloxycarbonyl groups, chloroacetyl groups or formyl groups. Free carboxyl groups may be derivatized to form salts, methyl and ethyl esters or other types of esters or 10 hydrazides. Free hydroxyl groups may be derivatized to form O-acyl or O-alkyl derivatives. The imidazole nitrogen of histidine may be derivatized to form N-imbenzylhistidine. Also included as chemical derivatives are those proteins or peptides which contain one or more naturally-occurring amino acid derivatives of the twenty standard amino acids. For examples: 4-hydroxyproline may be substituted for proline; 5-hydroxylysine may be 15 substituted for lysine; 3-methylhistidine may be substituted for histidine; homoserine may be substituted for serine; and ornithine may be substituted for lysine. The proteins or peptides of the present invention also include any protein or peptide having one or more additions and/or deletions or residues relative to the sequence of a peptide whose sequence is shown herein, so long as the peptide is biologically equivalent to the proteins or peptides of the 20 invention.

It is to be noted that, at the level of the amino acid sequence, at least one amino acids difference (with respect to known HCV amino acid sequences) is sufficient to be part of the invention, which means that the polypeptides of the invention correspond to polynucleic acids having at least one nucleotide difference (with known HCV polynucleic 25 acid sequences) involving an amino acid difference in the encoded polypeptide.

As the NS4 and the Core regions are known to contain several epitopes, for example characterized in patent application EP-A-0 489 968, and as the E1 protein is expected to be subject to immune attack as part of the viral envelope and expected to contain epitopes, the NS4, Core and E1 epitopes of the new types and subtypes disclosed 30 herein will consistently differ from the epitopes present in previously known genotypes. This is exemplified by the type-specificity of NS4 synthetic peptides as described in Simmonds et al. (1993c) and Stuyver et al. (1993b) and PCT/EP 94/01323 and the type-specificity of

recombinant E1 proteins as described in Maertens et al. (1994).

The peptides according to the present invention contain preferably at least 3, preferably 4, 5 contiguous HCV amino acids, 6, 7 preferably however at least 8 contiguous HCV amino acids, at least 10 or at least 15 (for instance at least 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50 or more amino acids).

TABLE 4

	Amino acids	Synonymous groups
10	Ser (S)	Ser, Thr, Gly, Asn
	Arg (R)	Arg, His, Lys, Glu, Gln
	Leu (L)	Leu; Ile, Met, Phe, Val, Tyr
	Pro (P)	Pro, Ala, Thr, Gly
15	Thr (T)	Thr, Pro, Ser, Ala, Gly, His, Gln
	Ala (A)	Ala, Pro, Gly, Thr
	Val (V)	Val, Met, Ile, Tyr, Phe, Leu, Val
	Gly (G)	Gly, Ala, Thr, Pro, Ser
	Ile (I)	Ile, Met, Leu, Phe, Val, Ile, Tyr
20	Phe (F)	Phe, Met, Tyr, Ile, Leu, Trp, Val
	Tyr (Y)	Tyr, Phe, Trp, Met, Ile, Val, Leu
	Cys (C)	Cys, Ser, Thr, Met
	His (H)	His, Gln, Arg, Lys, Glu, Thr
	Gln (Q)	Gln, Glu, His, Lys, Asn, Thr, Arg
25	Asn (N)	Asn, Asp, Ser, Gln
	Lys (K)	Lys, Arg, Glu, Gln, His
	Asp (D)	Asp, Asn, Glu, Gln
	Glu (E)	Glu, Gln, Asp, Lys, Asn, His, Arg
	Met (M)	Met, Ile, Leu, Phe, Val
30		

Table 4 Overview of the amino acid substitutions which could form the basis of analogs (muteins) as defined above

The polypeptides of the invention, and particularly the fragments, can be prepared

by classical chemical synthesis.

The synthesis can be carried out in homogeneous solution or in solid phase.

For instance, the synthesis technique in homogeneous solution which can be used is the one described by Houbenweyl in the book entitled "Methode der organischen chemie" (Method of organic chemistry) edited by E. Wunsh, vol. 15-I et II. THIEME, Stuttgart 1974.

The polypeptides of the invention can also be prepared in solid phase according to the methods described by Atherton and Shepard in their book entitled "Solid phase peptide synthesis" (IRL Press, Oxford, 1989).

The polypeptides according to this invention can be prepared by means of recombinant DNA techniques as described by Maniatis et al., Molecular Cloning: A Laboratory Manual, New York, Cold Spring Harbor Laboratory, 1982).

The present invention relates particularly to a polypeptide as defined above, comprising in its amino acid sequence at least one of the following amino acid residues:

I15, C38, V44, A49, Q43, P49, Q55, A58, S60 or D60, E68 or V68, H70, A71 or Q71 or N71, D72, H81, H101, D106, S110, L130, I134, E135, L140, S148, T150 or E150, Q153, F155, D157, G160, E165, I169, F181, L186, T190, T192 or I192 or H192, I193, A195, S196, R197 or N197 or K197, Q199 or D199 or H199, N199, F200 or T200, A208, I213, M216 or S216, N217 or S217 or G217 or K217, T218, I219, A222, Y223, I230, W231 or L231, S232 or H232 or A232, Q233, E235 or L235, F236 or T236, F237, L240 or M240, A242, N244, N249, I250 or K250 or R250, A252 or C252, A254, I255 or V255, D256 or M256, E257, E260 or K260, R261, V268, S272 or R272, I285, G290 or F290, A291, A293 or L293 or W293, T294 or A294, S295, H295, K296 or E296, Y297 or M297, I299 or Y299, I300, S301, P316, S2646, A2648, G2649, A2650, V2652, Q2653, H2656 or L2656, D2657, F2659, K2663 or Q2663, A2667 or V2667, D2677, L2681, M2686 or Q2686 or E2686, A2692 or K2692, H2697, I2707, L2708 or Y2708, A2709, A2719 or M2719, F2727, T2728 or D2728, E2729, F2730 or Y2730, I2741, I2745, V2746 or E2746 or L2746 or K2746, A2748, S2749 or P2749, R2750, E2751, D2752 or N2752 or S2752 or T2752 or V2752 or I2752 or Q2752, S2753 or D2753 or G2753, D2754, A2755, L2756 or Q2756, or R2757,

with said notation being composed of a letter representing the amino acid residue by its one-letter code, and a number representing the amino acid numbering according to Kato et al., 1990 as shown in Table 1 (see also the numbering in Figures 2, 4 and 6), or a part thereof which is unique to at least one of the HCV subtypes or types as defined in

Table 5, and which contains at least one amino acid differing from any of the known HCV types or subtypes, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide or part thereof.

These unique amino acid residues can be deduced from aligning the new HCV amino acid sequences as given in Figure 3 to all known HCV sequences. An alignment with the new sequences as represented in SEQ ID NO 1 to 106 is given in for instance Figures 2, 4 and 6. It should be clear that the alignments given in these figures may be completed with all known HCV sequences to illustrate that any of the above-given unique residues is indeed unique for at least one of the new HCV sequences of the present invention.

Within the group of unique and new amino acid residues of the present invention, unique residues may be found which are specific for the following new types (subtypes) of HCV according to the HCV classification system used in the present invention: type 1 subtype 1d, 1e, 1f or 1g isolates; type 2 subtype 2e, 2f, 2g, 2h, 2i, 2k or 2l isolates; type 3 subtype 3g isolates; type 4 subtype 4k, 4l or 4m isolates; type 7 subtype 7a, 7c or 7d isolates, type 9, type 10 or type 11 isolates. In order to obtain these residues the alignments given in Figures 2, 4 and 6 may be used to deduce the type- and or subtype-specificity of any of the unique residues given above.

For example T190 (detected in subtype 1d) refers to a threonine at position 190 (see Figure 2). In other sequences only a serine (S190) or exceptionally an alanine (A190 in type 10a) can be detected.

The polypeptides according to this embodiment of the invention may be possibly labelled, or attached to a solid substrate, or coupled to a carrier molecule such as biotin, or mixed with a proper adjuvant all known in the art and according to the intended use (diagnostic, therapeutic or prophylactic).

The present invention also relates to a polypeptide as defined above, comprising in its amino acid sequence at least one of the sequences represented by SEQ ID NO107 to 207 as listed above, or a part thereof which is unique to at least one of the HCV subtypes or types as defined in Table 5, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide or part thereof.

The present invention relates also to a polypeptide having an amino acid sequence as represented in any of SEQ ID NO 1 to 106, or a part thereof which is unique to at least one of the HCV subtypes or types as defined in Table 5, or an analog thereof being

substantially homologous and biologically equivalent to said polypeptide or part thereof.

The variable region in the core protein (V-CORE in Fig. 2) has been shown to be useful for serotyping (Machida et al., 1992). The sequence of the type 1 subtype 1d, 1e, 1f or 1g sequence; type 2 subtype 2e, 2f, 2g, 2h, 2i, 2k and 2l sequence; type 3 subtype 3g; type 4, subtype 4k, 4l or 4m sequence; type 7 (subtype 7a, 7c and 7d sequences), 9, 10 or 11 sequences of the present invention show type-specific features in this region. The peptide from amino acid 68 to 78 (V-core region) shows the following unique sequence for the sequences of the present invention (see figure 2):

ARQSDGRSWAQ or ARRSEGRSWAQ as for subtype 1d (SEQ ID NO 107 and  
10 108)  
ERRPEGRSWAQ as for subtype 1e (SEQ ID NO 109)  
ARRPEGRSWAQ as for subtype 1f (SEQ ID NO 110)  
DRRTTGKSWGR as for subtype 2k (SEQ ID NO 111)  
DRRATGRSWGR as for subtype 2e (SEQ ID NO 112)  
15 DRRATGKSWGR as for subtype 2f (SEQ ID NO 113)  
VRQPTGRSWGQ as for type 9 (SEQ ID NO 114)  
VRHQTGRTWAQ as for subtype 7a and 7c (SEQ ID NO 115)  
VRQNQGRTWAQ as for subtype 7d (SEQ ID NO 116)  
ARRTEGRSWAQ as for type 10 (SEQ ID NO 117)  
20 VRRTTGRXXXX or VRRTTGRTWAQ as for type 11 (SEQ ID NO 118 and  
119)

Five type-specific variable regions (V1 to V5) can be identified after aligning E1 amino acid sequences of the genotypes of the present invention to the genotypes already known, as shown in Figure 2.

25 Region V1 encompasses amino acids 192 to 203, this is the amino-terminal 10 amino acids of the E1 protein. The following unique sequences as shown in Fig. 2 can be deduced:

HEVRNASGVYHV or HEVRNASGVYHL as for subtype 1d, (SEQ ID NO 120  
and 121)  
30 YEVHSTTDGYHV as for subtype 1f (SEQ ID NO 122)  
VEVKNTSQAYMA as for subtype 2e (SEQ ID NO 123)  
IQVKNNSHFYMA as for subtype 2f (SEQ ID NO 124)

VQVKNTSTMYMA as for subtype 2g (SEQ ID NO 125)  
 VQVKNTSHSYMV as for subtype 2h (SEQ ID NO 126)  
 VQVANRSGSYMV as for subtype 2i (SEQ ID NO 127)  
 VEIKNTXNTYVL or VEIKNTSNTYVL as for subtype 2k (SEQ ID NO 128  
 and 129)

5

INYNRVSGIYYV or INYRNTSGIYHV or INYHNTSGIYHI or TNYRNVSGIYHV  
 as for subtype 4k (SEQ ID NO 130, 131, 132 or 133)

QHYRNVSGIYHV as for subtype 4l (SEQ ID NO 134)

IQVKNASGIYHL as for type 9 (SEQ ID NO 135)

10 AHYTNKSGLYHL as for subtype 7c (SEQ ID NO 136)

LNYANKSGLYHL as for subtype 7d (SEQ ID NO 137)

LEYRNASGLYMV as for type 10 (SEQ ID NO 138)

Region V2 encompasses amino acids 213 to 223. The following unique sequences  
 can be found in the V2 region as shown in Figure 2:

15 IYEMDGMIMHY or IYEMSGMILHA as for subtype 1d, (SEQ ID NO 139  
 and 140)

VYEAKDIILHT as for subtype 1f (SEQ ID NO 141)

VWQLXDAVLHV as for subtype 2e (SEQ ID NO 142)

VWQLRDAVLHV as for subtype 2f (SEQ ID NO 143)

20 IWQMKGAVLHV as for subtype 2g (SEQ ID NO 144)

VWQLKDAVLHV as for subtype 2h (SEQ ID NO 145)

VWQLEEAVLHV as for subtype 2i (SEQ ID NO 146)

TWQLXXAVLHV as for subtype 2k (SEQ ID NO 147)

VYEADHHILHL or VYEADHHILAL or VFEADHHILHL as for subtype 4k  
 (SEQ ID NO 148, 149 and 150)

25

VYESDHHILHL as for subtype 4l (SEQ ID NO 151)

VFEAETMILHL as for type 9 (SEQ ID NO 152)

VYEAETLILHL as for subtype 7c (SEQ ID NO 153)

VYEANGMILHL as for subtype 7d (SEQ ID NO 154)

30 VYEAGDIILHL as for type 10. (SEQ ID NO 155)

Region V3 encompasses the amino acids 230 to 242. The following unique V3  
 region sequences can be deduced from Figure 2:

VREDNHLRCWMAL or VRENNSSRCWMAL as for subtype 1d

(SEQ ID NO 156 and 157)

IREGNISRCWVLP as for subtype 1f (SEQ ID NO 158)

ENSSGRFHCWIPV as for subtype 2e (SEQ ID NO 159)

5 ERSGNRTFCWTAV as for subtype 2f (SEQ ID NO 160)

ELQGNKSRWCWIPV as for subtype 2g (SEQ ID NO 162)

ERHQNQSRWCWIPV as for subtype 2h (SEQ ID NO 163)

EWKDNTSRWCWIPV as for subtype 2i (SEQ ID NO 164)

EREGNSSRCWIPV as for subtype 2k (SEQ ID NO 165)

10 VREGNQSRCWVAL or VRTGNQSRCWVAL or VRVGNQSSCWVAL or  
VRVGNQSRCWVAL or VKEGNHSRCWVAL as for subtype 4k

(SEQ ID NO 166, 167, 168 or 169)

VKTGNTSRCWVAL as for subtype 4l (SEQ ID NO 170)

IKAGNESRCWLVP as for type 9 (SEQ ID NO 171)

15 VKXXNQSRCWVQA as for subtype 7c (SEQ ID NO 172)

VKTGNLTKCWLSA as for subtype 7d (SEQ ID NO 173)

VRSGNTSRWCWIPV as for type 10 (SEQ ID NO 174)

Region V4 encompasses the amino acids 248 to 257. The following unique V4

region sequences can be deduced from figure 2:

20 VKNASVPTAA or VKDANVPTAA as for subtype 1d (SEQ ID NO 175 and 176)

ARIANAPIDE as for subtype 1f (SEQ ID NO 177)

VSKPGALTKG as for subtype 2e (SEQ ID NO 178)

VSRPGALTRG as for subtype 2f (SEQ ID NO 179)

VNQPGALTRG as for subtype 2g (SEQ ID NO 180)

25 VSQPGALTRG as for subtype 2h (SEQ ID NO 181)

VSQPGALTKG as for subtype 2i (SEQ ID NO 182)

VSRPGALTEG as for subtype 2k (SEQ ID NO 183)

APYIGAPLES or APYTAAPLES as for subtype 4k (SEQ ID NO 184 and  
185)

30 APILSAPLMS as for subtype 4l (SEQ ID NO 186)

VPNSSVPIHG as for type 9 (SEQ ID NO 187)

VPNASTPVTG as for subtype 7c (SEQ ID NO 188)

VQNASVSIRG as for subtype 7d (SEQ ID NO 189)

VKSPCAATAS as for type 10 (SEQ ID NO 190)

Region V5 encompasses the amino acids 294 to 303. The following unique V5 region peptides can be deduced from figure 2:

5      SPRMHHTTQE or SPRLYHTTQE as for subtype 1d (SEQ ID NO 191 and 192)

TSRRHWTVQD as for subtype 1f (SEQ ID NO 193)

APKRHYFVQE as for subtype 2e (SEQ ID NO 194)

SPQYHTFVQE as for subtype 2f (SEQ ID NO 195)

10      SPQHHNFSQD as for subtype 2g (SEQ ID NO 196)

SPQHIFVQD as for subtype 2h (SEQ ID NO 197)

SPEHHHFVQD as for subtype 2k (SEQ ID NO 198)

RPRRHWTQD or RPRRHWTQD or QPRRHWTQD or RPRRHWTQD as for subtype 4k (SEQ ID NO 199, 200, 201 or 202)

15      QPRRHWTQD as for subtype 4l (SEQ ID NO 203)

RPKYHQVTQD as for type 9 (SEQ ID NO 204)

RPRMHQVVQE as for subtype 7c (SEQ ID NO 205)

RPRMYEIAQD as for subtype 7d (SEQ ID NO 206)

RHRQHWTVQD as for type 10 (SEQ ID NO 207)

20      The above given list of peptides are particularly useful for treatment and vaccine and diagnostic development.

Also comprised in the present invention is any synthetic peptide (see below) or polypeptide containing at least an epitope derived from the above-defined peptides in their peptidic chain. Also comprised within the present invention is any synthetic peptide or polypeptide comprising at least 6, 7, 8, or 9 contiguous amino acids derived from the above-defined peptides in their peptidic chain.

As used herein, 'epitope' or 'antigenic determinant' means an amino acid sequence that is immunoreactive. Generally an epitope consists of at least 3 to 4 amino acids, and more usually, consists of at least 5 or 6 amino acids, sometimes the epitope consists of about 7 to 8, or even about 10 amino acids.

The present invention particularly relates to any peptide (see below) or polypeptide contained in any of the amino acid sequences as represented in SEQ ID NO 2, 4, 7, 9, 12,

14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104 or 106 (see Table 5 and Figure 3, Examples section).

The present invention also relates to a recombinant polypeptide encoded by a polynucleic acid as defined above, or a part thereof which is unique to any of the HCV subtypes or types as defined in Table 5, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide.

The present invention also relates to a recombinant expression vector comprising a polynucleic acid or a part thereof as defined above, operably linked to prokaryotic, eukaryotic or viral transcription and translation control elements.

In general said recombinant vector will comprise a vector sequence, an appropriate prokaryotic, eukaryotic or viral promoter sequence followed by the nucleotide sequences as defined above, with said recombinant vector allowing the expression of any one of the polypeptides as defined above in a prokaryotic, or eukaryotic host or in living mammals when injected as naked DNA, and more particularly a recombinant vector allowing the expression of any of the new HCV sequences of the invention spanning particularly the following amino acid positions:

- a polypeptide starting in the region between positions 1 and 10 and ending at any position in the region between positions 70 and 420, more particularly a polypeptide spanning positions 1 to 70, 1 to 85, positions 1 to 120, positions 1 to 150, positions 1 to 191, or positions 1 to 200, for expression of the Core protein, and a polypeptide spanning positions 1 to 263, positions 1 to 326, positions 1 to 383, or positions 1 to 420 for expression of the Core and E1 protein;
- a polypeptide starting at any position in the region between positions 117 and 192, and ending at any position in the region between positions 263 and 420, for expression of E1, or forms that have the hydrophobic region deleted (positions 264 to 293 plus or minus 8 amino acids);
- a polypeptide starting at any position in the region between positions 1556 and 1688, and ending at any position in the region between positions 1739 and 1764, for expression of NS4, more particularly ;a polypeptide starting at position 1658 and ending at position 1711, for expression of NS4a antigen, and more particularly, a polypeptide starting at position 1712 and ending in the region

between positions 1743 and 1972 (for instance 1712-1743, 1712-1764, 1712-1782, 1712-1972, 1712-1782, 1712-1902), for expression of NS4b antigen or parts thereof.

Any other HCV vector construction known in the art may also be used for the  
5 recombinant polypeptides of the present invention.

Also any of the known purification methods for recombinant proteins may be used for the production of the recombinant polypeptides of the present invention, particularly the HCV recombinant polypeptide purification methods as disclosed in PCT/EP 95/03031 in name of Innogenetics N.V.

10 The term "vector" may comprise a plasmid, a cosmid, a phage, or a virus or a transgenic animal. Particularly useful for vaccine development may be BCG or adenoviral vectors, as well as avipox recombinant viruses.

The present invention also relates to a method for the production of a recombinant polypeptide as defined above, comprising:

- 15 - transformation of an appropriate cellular host with a recombinant vector, in which a polynucleic acid or a part thereof according to as defined above has been inserted under the control of appropriate regulatory elements,
- culturing said transformed cellular host under conditions enabling the expression of said insert, and,
- 20 - harvesting said polypeptide.

The term 'recombinantly expressed' used within the context of the present invention refers to the fact that the proteins of the present invention are produced by recombinant expression methods be it in prokaryotes, or lower or higher eukaryotes as discussed in detail below.

25 The term 'lower eukaryote' refers to host cells such as yeast, fungi and the like. Lower eukaryotes are generally (but not necessarily) unicellular. Preferred lower eukaryotes are yeasts, particularly species within Saccharomyces, Schizosaccharomyces, Kluveromyces, Pichia (e.g. Pichia pastoris), Hansenula (e.g. Hansenula polymorpha), Yarrowia, Schwaniomyces, Schizosaccharomyces, Zygosaccharomyces and the like.  
30 Saccharomyces cerevisiae, S. carlsbergensis and K. lactis are the most commonly used yeast hosts, and are convenient fungal hosts.

The term 'prokaryotes' refers to hosts such as E.coli, Lactobacillus, Lactococcus,

Salmonella, Streptococcus, Bacillus subtilis or Streptomyces. Also these hosts are contemplated within the present invention.

The term 'higher eukaryote' refers to host cells derived from higher animals, such as mammals, reptiles, insects, and the like. Presently preferred higher eukaryote host cells are derived from Chinese hamster (e.g. CHO), monkey (e.g. COS and Vero cells), baby hamster kidney (BHK), pig kidney (PK15), rabbit kidney 13 cells (RK13), the human osteosarcoma cell line 143 B, the human cell line HeLa and human hepatoma cell lines like Hep G2, and insect cell lines (e.g. Spodoptera frugiperda). The host cells may be provided in suspension or flask cultures, tissue cultures, organ cultures and the like. Alternatively the host cells may also be transgenic animals.

The term 'recombinant polynucleotide or nucleic acid' intends a polynucleotide or nucleic acid of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation : (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature, (2) is linked to a polynucleotide other than that to which it is linked in nature, or (3) does not occur in nature.

The term 'recombinant host cells', 'host cells', 'cells', 'cell lines', 'cell cultures', and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refer to cells which can be or have been, used as recipients for a recombinant vector or other transfer polynucleotide, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

The term 'replicon' is any genetic element, e.g., a plasmid, a chromosome, a virus, a cosmid, etc., that behaves as an autonomous unit of polynucleotide replication within a cell; i.e., capable of replication under its own control.

The term 'vector' is a replicon further comprising sequences providing replication and/or expression of a desired open reading frame.

The term 'control sequence' refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, splicing sites and terminators; in eukaryotes, generally, such control sequences include promoters,

splicing sites, terminators and, in some instances, enhancers. The term 'control sequences' is intended to include, at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences which govern secretion.

- 5       The term 'promoter' is a nucleotide sequence which is comprised of consensus sequences which allow the binding of RNA polymerase to the DNA template in a manner such that mRNA production initiates at the normal transcription initiation site for the adjacent structural gene.

The expression 'operably linked' refers to a juxtaposition wherein the components so  
10 described are in a relationship permitting them to function in their intended manner. A control sequence 'operably linked' to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The segment of the HCV cDNA encoding the desired sequence inserted into the  
15 vector sequence may be attached to a signal sequence. Said signal sequence may be that from a non-HCV source, e.g. the IgG or tissue plasminogen activator (tpa) leader sequence for expression in mammalian cells, or the  $\alpha$ -mating factor sequence for expression into yeast cells, but particularly preferred constructs according to the present invention contain signal sequences appearing in the HCV genome before the respective start points of the  
20 proteins.

A variety of vectors may be used to obtain recombinant expression of HCV single or specific oligomeric envelope proteins of the present invention. Lower eukaryotes such as yeasts and glycosylation mutant strains are typically transformed with plasmids, or are transformed with a recombinant virus. The vectors may replicate within the host  
25 independently, or may integrate into the host cell genome.

Higher eukaryotes may be transformed with vectors, or may be infected with a recombinant virus, for example a recombinant vaccinia virus. Techniques and vectors for the insertion of foreign DNA into vaccinia virus are well known in the art, and utilize, for example homologous recombination. A wide variety of viral promoter sequences, possibly  
30 terminator sequences and poly(A)-addition sequences, possibly enhancer sequences and possibly amplification sequences, all required for the mammalian expression, are available in the art. Vaccinia is particularly preferred since vaccinia halts the expression of host cell

Also known are insect expression transfer vectors derived from baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV), which is a helper-independent viral expression vector. Expression vectors derived from this system usually use the strong viral polyhedrin gene promoter to drive the expression of heterologous genes. Different vectors as well as methods for the introduction of heterologous DNA into the desired site of baculovirus are available to the man skilled in the art for baculovirus expression. Also different signals for posttranslational modification recognized by insect cells are known in the art.

15           The present invention also relates to a method for detecting antibodies to HCV present in a biological sample, comprising:

          (i) contacting the biological sample to be analysed for the presence of HCV with a polypeptide as defined above,

          (ii) detecting the immunological complex formed between said antibodies and said

20 polypeptide.

(i) contacting the biological sample to be analysed for the presence of HCV with a polypeptide as defined above,

(ii) detecting the immunological complex formed between said antibodies and said polypeptide.

The present invention also relates to a diagnostic kit for HCV typing, said kit comprising at least one polypeptide as defined above, with said polypeptide being preferably bound to a solid support.

The present invention also relates to diagnostic kit according as defined above, said

kit comprising a range of said polypeptides which are attached to specific locations on a solid substrate.

The present invention also relates to a diagnostic kit as defined above, wherein said solid support is a membrane strip and said polypeptides are coupled to the membrane in the form of parallel lines.

The immunoassay methods according to the present invention may utilize antigens from the different domains of the new and unique polypeptide sequences of the present invention that maintain linear (in case of peptides) and conformational epitopes (in case of polypeptides) recognized by antibodies in the sera from individuals infected with HCV. It is within the scope of the invention to use for instance single or specific oligomeric antigens, dimeric antigens, as well as combinations of single or specific oligomeric antigens. The HCV antigens of the present invention may be employed in virtually any assay format that employs a known antigen to detect antibodies. Of course, a format that denatures the HCV conformational epitope should be avoided or adapted. A common feature of all of these assays is that the antigen is contacted with the body component suspected of containing HCV antibodies under conditions that permit the antigen to bind to any such antibody present in the component. Such conditions will typically be physiologic temperature, pH and ionic strength using an excess of antigen. The incubation of the antigen with the specimen is followed by detection of immune complexes comprised of the antigen.

Design of the immunoassays is subject to a great deal of variation, and many formats are known in the art. Protocols may, for example, use solid supports, or immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, enzymatic, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the immune complex are also known; examples of which are assays which utilize biotin and avidin or streptavidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

The immunoassay may be, without limitation, in a heterogeneous or in a homogeneous format, and of a standard or competitive type. In a heterogeneous format, the polypeptide is typically bound to a solid matrix or support to facilitate separation of the sample from the polypeptide after incubation. Examples of solid supports that can be used are nitrocellulose (e.g., in membrane or microtiter well form), polyvinyl chloride (e.g., in sheets or microtiter wells), polystyrene latex (e.g., in beads or microtiter plates,

polyvinylidene fluoride (known as Immulon<sup>TM</sup>), diazotized paper, nylon membranes, activated beads, and Protein A beads. For example, Dynatech Immulon<sup>TM</sup> 1 or Immulon<sup>TM</sup> 2 microtiter plates or 0.25 inch polystyrene beads (Precision Plastic Ball) can be used in the heterogeneous format. The solid support containing the antigenic  
 5 polypeptides is typically washed after separating it from the test sample, and prior to detection of bound antibodies. Both standard and competitive formats are known in the art.

In a homogeneous format, the test sample is incubated with the combination of antigens in solution. For example, it may be under conditions that will precipitate any  
 10 antigen-antibody complexes which are formed. Both standard and competitive formats for these assays are known in the art.

In a standard format, the amount of HCV antibodies in the antibody-antigen complexes is directly monitored. This may be accomplished by determining whether labeled anti-xenogeneic (e.g. anti-human) antibodies which recognize an epitope on anti-HCV  
 15 antibodies will bind due to complex formation. In a competitive format, the amount of HCV antibodies in the sample is deduced by monitoring the competitive effect on the binding of a known amount of labeled antibody (or other competing ligand) in the complex.

Complexes formed comprising anti-HCV antibody (or in the case of competitive assays, the amount of competing antibody) are detected by any of a number of known  
 20 techniques, depending on the format. For example, unlabeled HCV antibodies in the complex may be detected using a conjugate of anti-xenogeneic Ig complexed with a label (e.g. an enzyme label).

In an immunoprecipitation or agglutination assay format the reaction between the HCV antigens and the antibody forms a network that precipitates from the solution or  
 25 suspension and forms a visible layer or film of precipitate. If no anti-HCV antibody is present in the test specimen, no visible precipitate is formed.

There currently exist three specific types of particle agglutination (PA) assays. These assays are used for the detection of antibodies to various antigens when coated to a support. One type of this assay is the hemagglutination assay using red blood cells (RBCs)  
 30 that are sensitized by passively adsorbing antigen (or antibody) to the RBC. The addition of specific antigen antibodies present in the body component, if any, causes the RBCs coated with the purified antigen to agglutinate.

To eliminate potential non-specific reactions in the hemagglutination assay, two artificial carriers may be used instead of RBC in the PA. The most common of these are latex particles. However, gelatin particles may also be used. The assays utilizing either of these carriers are based on passive agglutination of the particles coated with purified  
5 antigens.

The HCV antigens of the present invention comprised of conformational epitopes will typically be packaged in the form of a kit for use in these immunoassays. The kit will normally contain in separate containers the native HCV antigen, control antibody formulations (positive and/or negative), labeled antibody when the assay format requires the  
10 same and signal generating reagents (e.g. enzyme substrate) if the label does not generate a signal directly. The native HCV antigen may be already bound to a solid matrix or separate with reagents for binding it to the matrix. Instructions (e.g. written, tape, CD-ROM, etc.) for carrying out the assay usually will be included in the kit.

Immunoassays that utilize the native HCV antigen are useful in screening blood for  
15 the preparation of a supply from which potentially infective HCV is lacking. The method for the preparation of the blood supply comprises the following steps. Reacting a body component, preferably blood or a blood component, from the individual donating blood with HCV polypeptides of the present invention to allow an immunological reaction between HCV antibodies, if any, and the HCV antigen. Detecting whether anti-HCV antibody - HCV  
20 antigen complexes are formed as a result of the reacting. Blood contributed to the blood supply is from donors that do not exhibit antibodies to the native HCV antigens.

In cases of a positive reactivity to the HCV antigen, it is preferable to repeat the immunoassay to lessen the possibility of false positives. For example, in the large scale screening of blood for the production of blood products (e.g. blood transfusion, plasma,  
25 Factor VIII, immunoglobulin, etc.) 'screening' tests are typically formatted to increase sensitivity (to insure no contaminated blood passes) at the expense of specificity; i.e. the false-positive rate is increased. Thus, it is typical to only defer for further testing those donors who are 'repeatedly reactive'; i.e. positive in two or more runs of the immunoassay on the donated sample. However, for confirmation of HCV-positivity, the 'confirmation' tests  
30 are typically formatted to increase specificity (to insure that no false-positive samples are confirmed) at the expense of sensitivity.

The solid phase selected can include polymeric or glass beads, nitrocellulose,

2044020 "BETTER"

microparticles, microwells of a reaction tray, test tubes and magnetic beads. The signal generating compound can include an enzyme, a luminescent compound, a chromogen, a radioactive element and a chemiluminescent compound. Examples of enzymes include alkaline phosphatase, horseradish peroxidase and beta-galactosidase. Examples of  
 5 enhancer compounds include biotin, anti-biotin and avidin. Examples of enhancer compounds binding members include biotin, anti-biotin and avidin. In order to block the effects of rheumatoid factor-like substances, the test sample is subjected to conditions sufficient to block the effect of rheumatoid factor-like substances. These conditions comprise contacting the test sample with a quantity of anti-human IgG to form a mixture,  
 10 and incubating the mixture for a time and under conditions sufficient to form a reaction mixture product substantially free of rheumatoid factor-like substance.

The present invention particularly relates to an immunoassay format in which the polypeptides (or peptides) of the invention are coupled to a membrane in the form of parallel lines . This assay format is particularly advantageous for HCV typing purposes.

15 The present invention also relates to a pharmaceutical composition comprising at least one (recombinant) polypeptides as defined above and a suitable excipient, diluent or carrier.

The present invention also relates to a method of preventing HCV infection, comprising administering the pharmaceutical composition as defined above to a mammal in  
 20 effective amount to stimulate the production of protective antibody or protective T-cell response.

The present invention relates to the use of a composition as defined above in a method for preventing HCV infection.

The present invention further relates to a vaccine for immunizing a mammal against  
 25 HCV infection, comprising at least one (recombinant) polypeptide as defined above, in a pharmaceutically acceptable carrier.

The term 'immunogenic' refers to the ability of a substance to cause a humoral and/or cellular response, whether alone or when linked to a carrier, in the presence or absence of an adjuvant. 'Neutralization' refers to an immune response that blocks the  
 30 infectivity, either partially or fully, of an infectious agent. A 'vaccine' is an immunogenic composition capable of eliciting protection against HCV, whether partial or complete. A vaccine may also be useful for treatment of an individual, in which case it is called a

therapeutic vaccine.

The term 'therapeutic' refers to a composition capable of treating HCV infection.

The term 'effective amount' refers to an amount of epitope-bearing polypeptide sufficient to induce an immunogenic response in the individual to which it is administered, or  
 5 to otherwise detectably immunoreact in its intended system (e.g., immunoassay). Preferably, the effective amount is sufficient to effect treatment, as defined above. The exact amount necessary will vary according to the application. For vaccine applications or for the generation of polyclonal antiserum / antibodies, for example, the effective amount may vary depending on the species, age, and general condition of the individual, the  
 10 severity of the condition being treated, the particular polypeptide selected and its mode of administration, etc. It is also believed that effective amounts will be found within a relatively large, non-critical range. An appropriate effective amount can be readily determined using only routine experimentation. Preferred ranges of proteins for prophylaxis of HCV disease are 0.01 to 100 µg/dose, preferably 0.1 to 50 µg/dose. Several doses may be needed per  
 15 individual in order to achieve a sufficient immune response and subsequent protection against HCV disease.

The present invention also relates to a vaccine as defined above, comprising at least one (recombinant) polypeptide as defined above, with said polypeptide being unique for at least one of the subtypes or types as defined above.

20 Said vaccine compositions may include prophylactic as well as therapeutic vaccine compositions.

Pharmaceutically acceptable carriers include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins,  
 25 polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers; and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to : aluminim hydroxide (alum), N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-  
 30 MDP) as found in U.S. Patent No. 4,606,918, N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE) and RIBI, which contains three

components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate, and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion. Any of the 3 components MPL, TDM or CWS may also be used alone or combined 2 by 2. Additionally, adjuvants such as Stimulon (Cambridge Bioscience, Worcester, MA).

5 Immunogenic compositions used as vaccines comprise a 'sufficient amount' or 'an immunologically effective amount' of the proteins of the present invention, as well as any other of the above mentioned components, as needed. 'Immunologically effective amount', means that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment, as defined above. This amount varies depending  
 10 upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (e.g. nonhuman primate, primate, etc.), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, the strain of infecting HCV, and other relevant factors. It is expected that the amount will fall in a  
 15 relatively broad range that can be determined through routine trials. Usually, the amount will vary from 0.01 to 1000 µg/dose, more particularly from 0.1 to 100 µg/dose.

The proteins of the invention may also serve as vaccine carriers to present homologous (e.g. T cell epitopes or B cell epitopes from for instance the core, E1, E2, NS2, NS3, NS4 or NS5 regions) or heterologous (non-HCV) haptens, in the same manner as  
 20 Hepatitis B surface antigen (see European Patent Application 174,444). In this use, envelope proteins provide an immunogenic carrier capable of stimulating an immune response to haptens or antigens conjugated to the aggregate. The antigen may be conjugated either by conventional chemical methods, or may be cloned into the gene encoding E1 and/or E2 at a location corresponding to a hydrophilic region of the protein.  
 25 Such hydrophilic regions include the V1 region (encompassing amino acid positions 191 to 202), the V2 region (encompassing amino acid positions 213 to 223), the V3 region (encompassing amino acid positions 230 to 242), the V4 region (encompassing amino acid positions 230 to 242), the V5 region (encompassing amino acid positions 294 to 303) and the V6 region (encompassing amino acid positions 329 to 336). Another useful location for  
 30 insertion of haptens is the hydrophobic region (encompassing approximately amino acid positions 264 to 293). It is shown in the present invention that this region can be deleted without affecting the reactivity of the deleted E1 protein with antisera. Therefore, haptens

may be inserted at the site of the deletion.

The immunogenic compositions are conventionally administered parenterally, typically by injection, for example, subcutaneously or intramuscularly. Additional formulations suitable for other methods of administration include oral formulations and  
5 suppositories. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

The administration of the immunogen(s) of the present invention may be for either a prophylactic or therapeutic purpose. When provided prophylactically, the immunogen(s) is  
10 provided in advance of any exposure to HCV or in advance of any symptom of any symptoms due to HCV infection. The prophylactic administration of the immunogen serves to prevent or attenuate any subsequent infection of HCV in a mammal. When provided therapeutically, the immunogen(s) is provided at (or shortly after) the onset of the infection or at the onset of any symptom of infection or disease caused by HCV. The therapeutic  
15 administration of the immunogen(s) serves to attenuate the infection or disease.

In addition to use as a vaccine, the compositions can be used to prepare antibodies to HCV (E1) proteins. The antibodies can be used directly as antiviral agents. To prepare antibodies, a host animal is immunized using the E1 proteins native to the virus particle bound to a carrier as described above for vaccines. The host serum or plasma is collected  
20 following an appropriate time interval to provide a composition comprising antibodies reactive with the (E1) protein of the virus particle. The gamma globulin fraction or the IgG antibodies can be obtained, for example, by use of saturated ammonium sulfate or DEAE Sephadex, or other techniques known to those skilled in the art. The antibodies are substantially free of many of the adverse side effects which may be associated with other  
25 anti-viral agents such as drugs.

The present invention also relates particularly to a peptide corresponding to an amino acid sequence encoded by at least one of the HCV genomic sequences as defined above, with said peptide being unique to any of the HCV subtypes or types as defined in Table 5, and which contains at least one amino acid differing from any of the known HCV  
30 types or subtypes, or an analog thereof being substantially homologous and biologically equivalent.

The present invention relates particularly to a peptide comprising at least one unique

The present invention relates also particularly to a peptide comprising in its sequence a unique amino acid residue of the invention as defined above.

All the embodiments (immunoassay formats, vaccines, compositions, uses, etc.) illustrated for the polypeptides of the invention as above also relate to the peptides of the invention.

(i) contacting the biological sample to be analysed for the presence of HCV with a peptide as defined above,

The present invention also relates to a method for HCV typing, comprising:

(ii) detecting the immunological complex formed between said antibodies and said peptide.

The present invention also relates to a diagnostic kit for HCV typing, said kit comprising at least one peptide as defined above, with said peptide being preferably bound to a solid support.

- at least one NS4 peptide,
- at least one NS4 peptide and at least one Core peptide,
- at least one NS4 peptide and at least one Core peptide and at least one E1 peptide, - at least one NS4 peptide and at least one E1 peptide.

30        The present invention also relates to a diagnostic kit as defined above, said kit comprising a range of said peptides which are attached to specific locations on a solid substrate.

The present invention also relates to a diagnostic kit as defined above, wherein said solid support is a membrane strip and said peptides are coupled to the membrane in the form of parallel lines.

The present invention also relates to a pharmaceutical composition comprising at least one as defined above and a suitable excipient, diluent or carrier.

the present invention also relates to a method of preventing HCV infection, comprising administering the pharmaceutical composition as defined above to a mammal in effective amount to stimulate the production of protective antibody or protective T-cell response.

10 The present invention also relates to the use of a composition as defined above in a method for preventing HCV infection.

The present invention also relates to a vaccine for immunizing a mammal against HCV infection, comprising at least one peptide as defined above, in a pharmaceutically acceptable carrier.

15 The present invention relates also to a vaccine as defined above, comprising at least one peptide as defined above, with said peptide being unique for at least one of the subtypes or types as defined in Table 5.

The present invention relates to an antibody raised upon immunization with at least one polypeptide or peptide as defined above, with said antibody being specifically reactive 20 with any of said polypeptides or peptides, and with said antibody being preferably a monoclonal antibody.

The monoclonal antibodies of the invention can be produced by any hybridoma liable to be formed according to classical methods from splenic cells of an animal, particularly from a mouse or rat, immunized against the HCV polypeptides according to the 25 invention as defined above on the one hand, and of cells of a myeloma cell line on the other hand, and to be selected by the ability of the hybridoma to produce the monoclonal antibodies recognizing the polypeptides which has been initially used for the immunization of the animals.

The antibodies involved in the invention can be labelled by an appropriate label of 30 the enzymatic, fluorescent, or radioactive type.

The monoclonal antibodies according to this preferred embodiment of the invention may be humanized versions of mouse monoclonal antibodies made by means of

recombinant DNA technology, departing from parts of mouse and/or human genomic DNA sequences coding for H and L chains or from cDNA clones coding for H and L chains.

Alternatively the monoclonal antibodies according to this preferred embodiment of the invention may be human monoclonal antibodies. These antibodies according to the present embodiment of the invention can also be derived from human peripheral blood lymphocytes of patients infected with HCV type 1 subtype 1d, 1e, 1f or 1g, HCV type 2 subtype 2e, 2f, 2g, 2h, 2i, 2k or 2l; HCV type 3, subtype 3g; HCV type 4 subtype 4k, 4l or 4m; and/or HCV type 7 (subtypes 7a, 7c or 7d), 9, 10 or 11, or vaccinated against HCV. Such human monoclonal antibodies are prepared, for instance, by means of human peripheral blood lymphocytes (PBL) repopulation of severe combined immune deficiency (SCID) mice (for recent review, see Duchosal et al. 1992) or by screening Epstein Barr-virus-transformed lymphocytes of infected or vaccinated individuals for the presence of reactive B-cells by means of the antigens of the present invention.

The invention also relates to the use of the proteins of the invention, muteins thereof, or peptides derived therefrom for the selection of recombinant antibodies by the process of repertoire cloning (Persson et al., 1991).

Antibodies directed to peptides derived from a certain genotype may be used either for the detection of such HCV genotypes, or as therapeutic agents.

The present invention relates also to a method for detecting HCV antigens present in a biological sample, comprising:

- (i) contacting said biological sample with an antibody as defined above,
- (ii) detecting the immune complexes formed between said HCV antigens and said antibody.

The present invention relates also to a method for HCV typing, comprising:

- (i) contacting said biological sample with an antibody as defined above,
- (ii) detecting the immune complexes formed between said HCV antigens and said antibody.

The present invention relates also to a diagnostic kit for use in detecting the presence of HCV, said kit comprising at least one antibody as defined above, with said antibody being preferably bound to a solid support.

The present invention relates also to a diagnostic kit for HCV typing, said kit comprising at least one antibody as defined above, with said antibody being preferably

The present invention relates also to a diagnostic kit as defined above, said kit comprising a range of said antibodies which are attached to specific locations on a solid substrate.

The present invention relates also to a method of preventing or treating HCV infection, comprising administering the pharmaceutical composition as defined above to a mammal in effective amount.

The genotype may also be detected by means of a type-specific antibody as defined above, which may also linked to any polynucleotide sequence that can afterwards be amplified by PCR to detect the immune complex formed (Immuno-PCR, Sano et al., 1992).

15 Any publications or patent applications referred to herein are incorporated by reference. The following examples illustrate aspects of the invention but are in no way intended to limit the scope thereof.

**FIGURE LEGENDS****Figure Legends****Figure 1**

5

Alignment of the nucleotide sequences of the Core/E1 region of some of the isolates of the newly identified types and subtypes of the present invention, with other known prototype isolates of subtypes.

10 **Figure 2**

Alignment of the amino acid sequences of the Core/E1 region of some of the isolates of the newly identified types and subtypes of the present invention, with other known prototype isolates of subtypes.

15

**Figure 3**

Nucleotide and amino acid sequences obtained from the new HCV isolates of the present invention (SEQ ID NO 1 to 106).

20

**Figure 4**

Alignment of the amino acid sequences of the Core/E1 region of some of the isolates of the newly identified types and subtypes of the present invention, with other known prototype

25 isolates of subtypes.

**Figure 5**

Alignment of the nucleotide sequences of the NS5b region of some of the isolates of the  
30 newly identified types and subtypes of the present invention, with other known prototype  
isolates of subtypes.

**Figure 6**

5

Alignment of the amino acid sequences of the NS5b region of some of the isolates of the newly identified types and subtypes of the present invention, with other known prototype isolates of subtypes.

10 **Table 5**

Overview of the new subtypes and types of the present invention and the regions sequenced. The subtypes between brackets have been replaced by the non-bracketed subtypes following the classification of Tokita et al. (1994).

15

**Examples****Serum samples.**

Serum samples from Cameroonian blood donors (CAM) were screened for HCV  
20 antibodies with Innostest HCV Ab III, and confirmed by INNO-LIA HCV III (Innogenetics, Antwerp, Belgium). Serum samples from patients with chronic hepatitis C infection were obtained from various centers in the Benelux countries (BNL), from France (FR), from Pakistan (PAK), from Egypt (EG), and from Vietnam (VN).

Samples from the Benelux, Cameroon, France and Vietnam were selected because  
25 of their aberrant reactivities (isolates CAM1078, FR2, FR1, VN4, VN12, VN13, NE98 and others (see Table 5)).

**cPCR, LiPA, cloning and sequencing.**

30 RNA isolation, cDNA synthesis, PCR, cloning, and LiPA genotyping using biotinylated 5' UR amplification products were performed as described (Stuyver et al., 1994c). The 5' UR, the Core/E1, and the NS5B PCR products were used for direct

sequencing. The sequence of the universal 5' UR primers HCP95, HCP96, HCP98, and HCP29, were described previously (Stuyver et al. 1993b). The following primers were also described (Stuyver et al. 1994c): HCP41, a sense primer for the amplification of the Core region; HCP52 and HCP54 for amplification of the Core/E1 region; and HCP206 and  
 5 HCP207 for amplification of a 340-bp NS5B region.

Serum samples BNL1, BNL2, BNL3, BNL4, BNL5, BNL6, BNL7, BNL8, BNL9, BNL10, BNL11, BNL12, CAM1078, FR2, FR16, FR4, FR13, VN13, VN4, VN12, FR1, NE98, and FR19 were analyzed in the Core/E1 region by direct sequencing. Serum samples BNL1, BNL2, FR17, CAM1078, FR2, FR16, BNL3, FR4, BNL5, FR13, FR18, PAK64, BNL8,  
 10 BNL12, EG81, VN13, VN4, VN12, FR1, NE98, FR14, FR15, and FR19 were also analyzed in the NS5B region by direct sequencing. Partial 5' UR, Core, E1, and NS5B sequences were obtained. The length of the obtained sequences is sufficient to classify the obtained sequences into new types or subtypes, based on the phylogenetic distances to known sequences. The following sequences could be obtained (nucleotide sequences have odd-  
 15 numbered SEQ ID NO., amino acid sequences have even-numbered SEQ ID NO.): SEQ ID NO 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103 and 105. The amino acid sequences deduced therefrom are given in SEQ ID  
 20 NO 2, 4, 7, 9, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104 and 106. Table 5 gives an overview of these sequences.

Table 5

Type	Isolate	Nucleotide sequence position	
1d	BNL1	1-310 (SEQ ID NO. 1)	7932-8271 (SEQ ID NO. 53)
1d	BNL2	1-310 (SEQ ID NO. 5)	7932-8271 (SEQ ID NO. 55)
1d	FR17		7932-8271 (SEQ ID NO. 57)
1e	CAM1078	1-223 (SEQ ID NO. 9)	7932-8271 (SEQ ID NO. 61)
1f	FR2	1-950 (SEQ ID NO. 11)	7932-8271 (SEQ ID NO. 63)
1g	FR16	(-15)-816 (SEQ ID NO. 65)	7932-8271 (SEQ ID NO. 67)
2e	BNL3	1-310 (SEQ ID NO. 13)	7932-8271 (SEQ ID NO. 69)
2f	FR4	1-957 (SEQ ID NO. 17)	7932-8271 (SEQ ID NO. 71)
2g	BNL4		
2h	BNL5	1-310 (SEQ ID NO. 21)	7932-8271 (SEQ ID NO. 73)
2i	BNL6		
2k	FR13	(-238)-957 (SEQ ID NO. 75)	
2l	FR18		
3g	PAK64		
4k	BNL7	1-310 (SEQ ID NO. 27)	7932-8271 (SEQ ID NO. 83)
4k	BNL8		
4k	BNL9		
4k	BNL10		
4k	BNL11		
4l	BNL12		
4m	EG81		
7a(8b) VN13		1-413 (SEQ ID NO. 45)	7932-8271 (SEQ ID NO. 85)
7c(8a) VN4		1-957 (SEQ ID NO. 43)	7932-8271 (SEQ ID NO. 87)
7d(9a) VN12		1-957 (SEQ ID NO. 47)	7932-8271 (SEQ ID NO. 89)
9a(7a) FR1		1-957 (SEQ ID NO. 41)	7932-8271 (SEQ ID NO. 91)
10a	NE98	1-310 (SEQ ID NO. 49)	7932-8271 (SEQ ID NO. 93)
11a	FR14		7932-8271 (SEQ ID NO. 95)
11a	FR15		7932-8271 (SEQ ID NO. 97)
11a	FR19		7932-8266 (SEQ ID NO. 99)
		478-925 (SEQ ID NO. 51)	7932-8271 (SEQ ID NO. 101)
		(-238)-223 (SEQ ID NO. 103)	7932-8271 (SEQ ID NO. 105)
		478-925 (SEQ ID NO. 3)	
		478-925 (SEQ ID NO. 7)	
		(-238)-414 (SEQ ID NO. 59)	
		478-957 (SEQ ID NO. 15)	
		478-925 (SEQ ID NO. 19)	
		478-925 (SEQ ID NO. 23)	
		478-833 (SEQ ID NO. 25)	
		478-925 (SEQ ID NO. 29)	
		478-925 (SEQ ID NO. 31)	
		478-925 (SEQ ID NO. 33)	
		478-925 (SEQ ID NO. 35)	
		478-925 (SEQ ID NO. 37)	
		478-925 (SEQ ID NO. 39)	



### Phylogenetic analysis.

Previously published sequences were taken from the EMBL/Genbank database. Alignments were created using the program HCVALIGN (Stuyver et al. 1994c). Sequences were presented in a sequential format to the Phylogeny Inference Package (PHYLP) version 3.5c (public domain program freely available from the University of Washington, Seattle, USA). Distance matrices were produced by DNADIST using the Kimura 2-parameter setting and further analyzed in NEIGHBOR, using the neighbor-joining setting. The program DRAWTREE was used to create graphic outputs.

#### 10 Identification of new subtypes

These analyses indicated the clustering of BNL1, BNL2, CAM 1078, FR2, FR16, and FR17 with type 1 isolates, yet neither of these sequences clustered together with any of the known type 1 subtypes 1a, 1b, or 1c. BNL1, BNL2, and FR17 clearly clustered together and could be assigned a new type 1 subtype 1d, while CAM1078 could be classified into another new subtype 1e, FR2 could be classified into another type 1 subtype 1f, and FR16 could be classified into yet another type 1 subtype 1g. Interestingly, all 3 type 1d isolates (BNL1, BNL2, and FR17) and 1g isolate FR16 were obtained from patients of Moroccan ethnic origin who resided in Europe.

Another group of isolates showed homology to other type 2 sequences, but none of the isolates BNL3, FR4, BNL4, BNL5, BNL6, FR13, or FR18 could be classified into one of the known type 2 subtypes 2a, 2b, 2c (Bukh et al., 1993), or 2d (Stuyver et al., 1994c). Based on the phylogenetic distances to other type 2 isolates and to other isolates of the group, each of these isolates could be classified into a new type 2 subtype. BNL3 was assigned subtype 2e, FR4 subtype 2f, BNL4 subtype 2g, BNL5 subtype 2h, and BNL6 could be classified into yet another type 2 subtype 2i. If the previously published isolate HN4 is classified as 2j, FR13 and FR18 may be classified into new type 2 subtypes 2k and 2l. However, the possibility that FR13 and FR18 could belong to subtypes 2g or 2i has not yet been ruled out. Definite classification can be obtained by determining the NS5B sequences of isolates BNL4 and BNL6, belonging to subtypes 2g and 2i, respectively.

30 Isolate PAK64 showed homology to type 3 sequences, but could not be classified into one of the known type 3 subtypes 3a to f. Based on the phylogenetic distances to other

type 3 isolates, PAK64 could be classified into a new type 3 subtype. PAK64 was assigned subtype 3g. However, the possibility that PAK64 belongs to a known type 3 subtype can not be strictly ruled out since only one region of the genome has been sequenced. Definite classification can be obtained by determining the Core/E1 sequences of isolate PAK64 after  
 5 amplification with primer HcPr52 and HcPr54.

Among the Benelux and Egyptian samples that were analyzed, some sequences clustered with the previously identified type 4 subtypes 4c and 4d. However, BNL7, BNL8, BNL9, BNL10, BNL11, BNL12, and EG81 clustered into new subtypes of type 4. Isolates BNL7, BNL8, BNL9, BNL10, and BNL11 clustered again separately from BNL12 and EG81  
 10 into a new subtype 4k. This subtype was the predominant subtype in the Benelux countries. BNL12 and EG81 also segregated into separate subtypes. BNL12 was assigned to another new subtype 4l and EG81 was assigned to yet another new subtype 4m.

#### 15 Identification of new HCV major types

Isolates FR1, VN4, VN12, VN13, NE98, FR14, FR15, and FR19 did not cluster with any of the known 6 major types of HCV. VN4, VN12, and VN13 were very distantly related to genotype 6, but phylogenetic analysis indicated that these isolates should be assigned new major types. VN13, VN4 and VN12 were related at the subtype level and assigned type  
 20 7a, 7c, and 7d, respectively. FR1 was not related to any known isolate and was assigned genotype 9a. NE98 shows a distant relatedness to type 3 sequences, yet phylogenetic analysis suggested classification into a new major type 10a. Depending on international guidelines for assigning type and subtype levels, NE98 may also be classified into an additional type 3 subtype. FR14, FR15, and FR19 show a very distant relatedness to type 2  
 25 sequences, yet phylogenetic analysis indicated these isolates to be classified into a new major type 11, all belonging to the same subtype designated 11a. Depending on international guidelines for assigning type and subtype levels, FR14, FR15, and FR19 may also be classified into an additional type 2 subtype.

## REFERENCES

- Barany F (1991). Genetic disease detection and DNA amplification using cloned  
5 thermostable ligase. *Proc Natl Acad Sci USA* 88: 189-193.
- Bej A, Mahbubani M, Miller R, Di Cesare J, Haff L, Atlas R (1990) Mutiplex PCR  
amplification and immobilized capture probes for detection of bacterial pathogens and  
indicators in water. *Mol Cell Probes* 4:353-365.
- 10 Bukh J, Purcell R, Miller R (1992). Sequence analysis of the 5' noncoding region of hepatitis  
C virus. *Proc Natl Acad Sci USA* 89:4942-4946.
- Bukh J, Purcell R, Miller R (1993). At least 12 genotypes of hepatitis C virus predicted by  
15 sequence analysis of the putative E1 gene of isolates collected worldwide. *Proc. Natl. Acad.  
Sci. USA* 90,8234-8238.
- Cha T, Beal E, Irvine B, Kolberg J, Chien D, Kuo G, Urdea M (1992) At least five related,  
but distinct, hepatitis C viral genotypes exist. *Proc Natl Acad Sci USA* 89:7144-7148.
- 20 Chan S-W, Simmonds P, McOmish F, Yap P, Mitchell R, Dow B, Follett E (1991)  
Serological responses to infection with three different types of hepatitis C virus. *Lancet*  
338:1991.
- 25 Chan S-W, McOmish F, Holmes E, Dow B, Peutherer J, Follett E, Yap P, Simmonds P  
(1992) Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing  
variants. *J Gen Virol* 73:1131-1141.
- Chomczynski P, Sacchi N (1987) Single step method of RNA isolation by acid guanidinium  
30 thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156-159.

Choo Q, Richman K, Han J, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina-Selby A, Barr P, Weiner A, Bradley D, Kuo G, Houghton M (1991) Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci USA* 88:2451-2455.

5 Compton J (1991). Nucleic acid sequence-based amplification. *Nature*, 350: 91-92.

Duchosal A, Eming S, Fisher P (1992) Immunization of hu-PBL-SCID mice and the rescue of human monoclonal Fab fragments through combinatorial libraries. *Nature* 355:258-262.

10 Duck P (1990). Probe amplifier system based on chimeric cycling oligonucleotides. *Biotechniques* 9, 142-147.

Guatelli J, Whitfield K, Kwoh D, Barringer K, Richman D, Gengeras T (1990) Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral  
15 replication. *Proc Natl Acad Sci USA* 87: 1874-1878.

Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Shimotohmo K (1991) Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. *Proc Natl Acad Sci USA* 88, 5547-5551.

20

Jacobs K, Rudersdorf R, Neill S, Dougherty J, Brown E, Fritsch E (1988) The thermal stability of oligonucleotide duplexes is sequence independent in tetraalkylammonium salt solutions: application to identifying recombinant DNA clones. *Nucl Acids Res* 16:4637-4650.

25 Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, Shimotohno K (1990) Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc Natl Acad Sci USA* 87:9524-9528.

Kwoh D, Davis G, Whitfield K, Chappelle H, Dimichele L, Gingeras T (1989). Transcription-  
30 based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format. *Proc Natl Acad Sci USA*, 86: 1173-

1177.

Kwok S, Kellogg D, McKinney N, Spasic D, Goda L, Levenson C, Sinisky J, (1990). Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nucl. Acids Res., 18: 999.

Landgren U, Kaiser R, Sanders J, Hood L (1988). A ligase-mediated gene detection technique. Science 241:1077-1080.

10 Lizardi P, Guerra C, Lomeli H, Tussie-Luna I, Kramer F (1988) Exponential amplification of recombinant RNA hybridization probes. Bio/Technology 6:1197-1202.

Lomeli H, Tyagi S, Printchard C, Lizardi P, Kramer F (1989) Quantitative assays based on the use of replicatable hybridization probes. Clin Chem 35: 1826-1831.

15

Machida A, Ohnuma H, Tsuda F, Munekata E, Tanaka T, Akahane Y, Okamoto H, Mishiro S (1992) Hepatology 16, 886-891.

Maniatis T, Fritsch E, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

20

Mori S, Kato N, Yagyu A, Tanaka T, Ikeda Y, Petchclai B, Chiewsilp P, Kurimura T, Shimotohno K (1992) A new type of hepatitis C virus in patients in Thailand. Biochem Biophys Res Comm 183:334-342.

25

Okamoto H, Okada S, Sugiyama Y, Kurai K, Iizuka H, Machida A, Miyakawa Y, Mayumi M (1991) Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. J Gen Virol 72:2697-2704.

30

Okamoto H, Kurai K, Okada S, Yamamoto K, Iizuka H, Tanaka T, Fukuda S, Tsuda F,

Mishiro S (1992) Full-length sequences of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. *Virology* 188:331-341.

Persson M, Caothien R, Burton D (1991). Generation of diverse high-affinity human  
5 monoclonal antibodies by repertoire cloning. *Proc Natl Acad Sci USA* 89:2432-2436.

Saiki R, Gelfand D, Stoffel S, Scharf S, Higuchi R, Horn G, Mullis K, Erlich H (1988).  
Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.  
*Science* 239:487-491.

10 Saiki R, Walsh P, Levenson C, Erlich H (1989) Genetic analysis of amplified DNA with  
immobilized sequence-specific oligonucleotide probes (1989) *Proc Natl Acad Sci USA*  
86:6230-6234.

Sano T, Smith C, Cantor C (1992) Immuno-PCR: very sensitive antigen detection by means  
15 of specific antibody-DNA conjugates. *Science* 258:120-122.

Simmonds P, McOmsh F, Yap P, Chan S, Lin C, Dusheiko G, Saeed A, Holmes E (1993a),  
Sequence variability in the 5' non-coding region of hepatitis C virus : identification of a new  
virus type and restrictions on sequence diversity. *J Gen Virology*, 74:661-668.

20

Stuyver L, Rossau R, Wyseur A, Duhamel M, Vanderborght B, Van Heuverswyn H,  
Maertens G (1993b) Typing of hepatitis C virus (HCV) isolates and characterization of new  
(sub)types using a Line Probe Assay. *J Gen Virology*, 74: 1093-1102.

25 Tokita et al. (1994) Hepatitis C virus variants from Vietnam are classifiable into the seventh,  
eighth, and ninth major genetic groups. *Proc. Natl. Acad. Sci*, 91: 11022-11026.

Walker G, Little M, Nadeau J, Shank D (1992). Isothermal in vitro amplification of DNA by a  
restriction enzyme/DNA polymerase system. *Proc Natl Acad Sci USA* 89:392-396.

30

Wu D, Wallace B (1989). The ligation amplification reaction (LAR) - amplification of specific

DNA sequences using sequential rounds of template-dependent ligation. *Genomics* 4:560-569.

Miller P, Yano J, Yano E, Carroll C, Jayaram K, Ts'o P (1979) Nonionic nucleic acid analogues. Synthesis and characterization of dideoxyribonucleoside methylphosphonates. *Biochemistry* 18(23):5134-43.

Nielsen P, Egholm M, Berg R, Buchardt O (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 254(5037):1497-500.

10

Nielsen P, Egholm M, Berg R, Buchardt O (1993) Sequence specific inhibition of DNA restriction enzyme cleavage by PNA. *Nucleic-Acids-Res.* 21(2):197-200.

Asseline U, Delarue M, Lancelot G, Toulme F, Thuong N (1984) Nucleic acid-binding molecules with high affinity and base sequence specificity : intercalating agents covalently linked to oligodeoxynucleotides. *Proc. Natl. Acad. Sci. USA* 81(11):3297-301.

Matsukura M, Shinozuka K, Zon G, Mitsuya H, Reitz M, Cohen J, Broder S (1987) Phosphorothioate analogs of oligodeoxynucleotides : inhibitors of replication and cytopathic effects of human immunodeficiency virus. *Proc. Natl. Acad. Sci. USA* 84(21):7706-10.

Maertens, G., Ducatteeuw, A., Stuyver, L., Vandeponseele, P., Venneman, A., Wyseur, A., Bosman, F., Heijtkink, R. & de Martynoff, G. (1994) Low prevalence of anti-E1 antibodies reactive to recombinant type 1b E1 envelope protein in type 2, 3, and 4 sera, but high prevalence in subtypes 1a and 1b. In: *Viral Hepatitis and Liver Disease,, Proceedings of the International Symposium on Viral Hepatitis and Liver Disease* (Eds. Nishioka, K., Suzuki, H., Mishiro, S., and Oda, T.), pp 314-316, Springer-Verlag Tokyo.

Simmonds, P., Rose, K.A., Graham, S., Chan, S.-W., McOmish, F., Dow, B.C., Follett, E.A.C., Yap, P.L., & Marsden, H. (1993b) Mapping of serotype-specific, immunodominant epitopes in the NS4 region of hepatitis C virus (HCV): Use of type-specific peptides to

serologically discriminate infections with HCV type 1, 2, and 3. *J. Clin. Microbiol.* **31**, 1493-1503.

Simmonds, P., Holmes, E.C., Cha, T.-A., Chan, S.-W., McOmish, F., Irvine, B., Beall, E.,  
 5 Yap, P.L., Kolberg, J., & Urdea, M.S. (1993c) *J. Gen. Virol.* **74**, 2391-2399.

Stuyver, L., Van Arnhem, W., Wyseur, A. & Maertens, G. (1994) Cloning and phylogenetic analysis of the Core, E2, and NS3/4 regions of hepatitis C virus type 5a. *Biochem. Biophys. Res. Comm.* **202**, 1308-1314.

10

Simmonds, P., Alberti, A., Alter, H., Bonino, F., Bradley, D.W., Br      , C., Brouwer, J.,  
 Chan, S.-W., Chayama K., Chen, D.-S., Choo, Q.-L., Colombo, M., Cuypers, T., Date, T.,  
 Dusheiko, G., Esteban, J.I., Fay, O., Hadziyannis, S., Han, J., Hatzakis, A., Holmes, E.C.,  
 Hotta, H., Houghton, M., Irvine, B., Kohara, M., Kolberg, J.A., Kuo, G., Lau, J.Y.N., Lelie,  
 15 P.N., Maertens, G., McOmish, F., Miyamura, T., Mizokami, M., Nomoto, A., Prince A.M.,  
 Reesink, H.W., Rice, C., Roggendorf, M., Schalm, S., Shikata, T., Shimotohno, K., Stuyver,  
L., Tr    , C., Weiner, A., Yap, P.L. & Urdea, M.S. (1994) A proposed system for the  
 nomenclature of hepatitis C virus genotypes. *Hepatology* **19**, 1321-1324.

20 Stuyver, L., Van Arnhem, W., Wyseur, A., DeLeys, R. & Maertens, G. (1993a) Analysis of  
 the putative E1 envelope and NS4a epitope regions of HCV type 3. *Biochem. Biophys. Res.*  
*Comm.* **192**, 635-641.

Stuyver, L., Rossau, R., Wyseur, A., Duhamel, M., Vanderborght, B., Van Heuverswyn, H. &  
 25 Maertens, G. (1993b) Typing of hepatitis C virus isolates and characterization of new  
 subtypes using a line probe assay. *J. Gen Virol.* **74**, 1093-1102.

Stuyver, L., Wyseur, A., Van Arnhem, W., Rossau, R., Delaporte, E., Dazza, M.-C., Van  
 Doorn, L.-J., Kleter, B. & Maertens, G. (1994a) The use of a line probe assay as a tool to  
 30 detect new types or subtypes of hepatitis C virus. In: *Viral Hepatitis and Liver Disease,*  
*Proceedings of the International Symposium on Viral Hepatitis and Liver Disease* (Eds.

Nishioka, K., Suzuki, H., Mishiro, S., and Oda, T.), pp 317-319, Springer-Verlag Tokyo.

Stuyver, L., Van Arnhem, W., Wyseur, A. & Maertens, G. (1994b) Cloning and Phylogenetic analysis of the Core, E2, and NS3/4 regions of the hepatitis C virus type 5a. Biochem. 5 Biophys. Res. Comm. **202**, 1308-1314.

Stuyver, L., Van Arnhem, W., Wyseur, A., Hernandez, F., Delaporte, E., & Maertens, G. (1994c) Classification of hepatitis C viruses based on phylogenetics analysis of the E1 and NS5B regions and identification of 5 new subtypes. Proc. Natl. Acad. Sci. USA **91**.

10

Stuyver et al. (1995) Hepatitis C virus genotyping by means of 5'-UR/core line probe assays and molecular analysis of untypeable samples. Virus Reasearch (in press).